Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Wendi Liang Date

**Analysis of Socioeconomic Disparities in Receipt of Adjuvant Chemotherapy in Stage II Colon Cancer: An Inquiry Based on National Cancer Database**

By

Wendi Liang

Master of Science in Public Health

Biostatistics

_____

Yuan Liu, Ph.D.

Committee Chair

_____

Shaffer, Virginia Oliva, M.D.

Committee Member

**Analysis of Socioeconomic Disparities in Receipt of Adjuvant Chemotherapy in Stage II Colon Cancer : An Inquiry Based on National Cancer Database**

By

Wendi Liang

B.Eng.

Hefei University of Technology

2014

Thesis Committee Chair: Yuan Liu, Ph.D.

An abstract of

A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Science in Public Health

in Biostatistics

2016

## Abstract

**Analysis of Socioeconomic Disparities in Receipt of Adjuvant Chemotherapy in Stage II Colon Cancer: An Inquiry Based on National Cancer Database**

By Wendi Liang

This study aimed to first assess whether socioeconomic covariates were predictors of receipt of chemotherapy in stage II colon cancer. Second, for patients who received adjuvant chemotherapy, this study investigated whether there exist predictors for overall survival. As a retrospective study of stage II colon cancer patients in the National Cancer Database, univariate analyses based on ANOVA and Chi-square test and Multivariate analyses using logistic regression and Cox regression models were carried out. Propensity score matching also was adopted in this study to further reduce the selection bias to assess the efficacy of adjuvant chemotherapy. The results demonstrated that socioeconomic factors including age, race, gender, insurance type, facility type, facility location, great circle distance, income and education level were significantly associated with the receipt of adjuvant chemotherapy in stage II colon cancer patients. Other clinical risk factors including tumor size, surgical margins, sequence number, and Charlson-Deyo score were also significantly influencing the receipt of chemotherapy. For patients received postoperative adjuvant chemotherapy, this analysis indicated that socioeconomic factors, such as race, sex, insurance type, living in communities with different income and education attainment were highly associated with the overall survival. Besides, this study confirmed the significant overall survival benefit of chemotherapy in stage II colon cancer.

# Analysis of Socioeconomic Disparities in Receipt of Adjuvant Chemotherapy in Stage II Colon Cancer: An Inquiry Based on National Cancer Database

By

Wendi Liang

B.Eng.

Hefei University of Technology

2014

Thesis Committee Chair: Yuan Liu, Ph.D.

A thesis submitted to the Faculty of the

Rollins School of Public Health of Emory University

in partial fulfillment of the requirements for the degree of
Master of Science in Public Health

in Biostatistics

2016

**Acknowledgements**

First, I am extremely grateful for my thesis advisor, Dr. Yuan Liu of the Biostatistics and Bioinformatics Department at Emory University, whose inspiration, guidance, and support made this thesis possible. I enjoyed working under Dr. Liu's guidance, and learned a lot through this study.

In addition, I am thankful to Dr. Shaffer, Virginia Oliva for her support and time. Also, I wish to thank the principle investigator Dr. Gillespie, Theresa Wicklin, for her advices on this project.

Furthermore, it is a great honor for me to pursue my Master of Science in Public Health degree in the Department of Biostatistics and Bioinformatics at Emory University. This valuable experience is a great help to my future career.

Finally, I would like to thank my parents, for their encouragement and support throughout my life and education.

## Table of Contents

## List of Tables

## List of Figures

# 1. Introduction

As the mortality of colon cancer in the United States has declined over the past three decades (Haggar & Boushey, 2009; Sharma & O'Keefe, 2007), colorectal cancer has remained a top reason of mortality throughout the world (Organization, 2002). Chemotherapy become widely used in colon cancer, especially for stage III patients (Group, 2007). However, the recommendation and usage of chemotherapy for stage II colon cancer is not very clear, regarding the toxicity, inconvenience of treatment, and cost (Benson et al., 2004). Although there exist evidence for the survival benefit of adjuvant chemotherapy in stage II colon cancer (Group, 2007; McKenzie et al., 2011), whether socioeconomic factors could be used to predict the disparity in usage of chemotherapy is unclear. Therefore, there is a need to further confirm the survival benefit of chemotherapy for stage II colon cancer, as well as identify socioeconomic factors, which may associated with the disparity in chemotherapy usage.

To address the above uncertainty, this study first assessed whether the socioeconomic covariates were predictors in receipt of chemotherapy (Aim1) and potential benefit of chemotherapy in terms of overall survival. Then for patients received adjuvant chemotherapy, this study investigated whether these covariates were predictors for overall survival (Aim2). This retrospective study was data from the National Cancer Data Base (NCDB) of stage II colon cancer patients diagnosed between 2004 and 2012. Furthermore, to make the two groups of patients more comparable, an evaluation of survival benefit for chemotherapy based on propensity score matching were provided.

The confirmation of the benefit of adjuvant chemotherapy would provide helpful information for both patients and physicians on the usage of chemotherapy, and let them make a better decision (Group, 2007). Besides, by answering these questions, we would be clearer with the covariates that lead to proceed chemotherapy. At the same time, the association between the covariates and a better after-Chemo survival would be evaluated. This study aimed to help provide insight to facilitate better care for patients in stage II colon cancer. The useful information may also help with making optimal decisions of receipt chemotherapy leading to better clinical outcomes.

# 2. Background

The following background review provided necessary context to understand the aims of the study. The review first discusses the prevalence of colon cancer with the corresponding standard recommend treatments. Second, the review offers an overview of adjuvant chemotherapy usage and efficacy, with the necessity for enhanced understanding the benefit of implementing adjuvant chemotherapy for stage II colon cancer. Finally, the review elucidated the factors that potentially affect the receipt of adjuvant chemotherapy, particularly socioeconomic factors.

## Colon cancer and standard treatments

Colorectal cancer is considered to be the third most common cancer worldwide, with estimated 1,023,256 new cases and 529,020 deaths per year (Kamangar, Dores, & Anderson, 2006). The prevalence is also high in the United States, with approximately 106,000 new diagnoses and over 50,000 deaths per year (McKenzie et al., 2011). While

underwent surgery to remove the cancer is widely used for early-stage colon cancer, chemotherapy is the main postoperative treatment for patients with advanced colon cancer (DeSantis et al., 2014). Oxaliplatin, recently combined with the fluorouracil has demonstrated significantly reduce the recurrence rate and improved the survival rate for patients with resected colon cancer (Dienstmann, Salazar, & Tabernero, 2015). Cytotoxic chemotherapy, can also lower the risk of tumor recurrence for patients with resection, without a clear benefit and regimens (Group, 2007).

## Adjuvant chemotherapy

To improve survival and reduce recurrence in colon cancer, the Fluorouracil-based adjuvant chemotherapy has been used as the standard adjuvant chemotherapy treatment (André et al., 2004; Sharlene Gill et al., 2004). Based on pharmacological rationale, adjuvant fluorouracil and folinic acid were confirmed to be benefit (Kerr, 1989). Common chemotherapy treatment will perform approximately 6 months after surgery, particular for stage III and some stage II colon cancer patients. (DeSantis et al., 2014).

Although it is recommended for patients with stage III colon cancer to receive postoperative chemotherapy treatment, the benefit of adjuvant chemotherapy for stage II patients remains under dispute (Benson et al., 2004; Kerr, 1989; McKenzie et al., 2011). This debate was sustained for a long period by the contradictory ideas of two groups of researcher (André et al., 2004). While international Multicentre Pooled Analysis of B2 Colon Cancer Trials (IMPACT B2) claimed that no statistical significant benefit was detected for stage II tumors, (Mamounas, Bear, Atkins, & Song, 1999), the QUASAR trial, which aimed to evaluate any survival benefit from adjuvant chemotherapy for

colorectal cancer, concluded that chemotherapy have survival benefit of stage II patients (Group, 2007). However, the Quasar mentioned that the absolute improvements were small (RR: 0.82 with 95% CI: 0.7 to 0.95), while the National Surgical Adjuvant Breast and Bowel Project pointed that the benefits of treatment were equal for stage II and stage III tumors (Group, 2007; E. Mamounas et al., 1999). Current recommendations for stage II disease is that for patients with high risk features chemotherapy should be offered (Benson et al., 2004). Also, a recent meta-analysis, which evaluated 3300 patients enrolled in randomized trials, claimed that patients with stage II disease gain benefit through adjuvant chemotherapy (Sharlene Gill et al., 2004). The overall survival at 5 years in Dr. Gill's study improved from 64% to 71% for patients with adjuvant chemotherapy. (S. Gill et al., 2004).

## Factors affect receipt of adjuvant chemotherapy

While there exist suggestions that age at diagnosis, rural–urban disparities and clinical factors, such as lymph node status influencing receipt of adjuvant chemotherapy among breast cancer patients (Zhang, Gao, Bu, Fan, & Jia, 2013), age had also be treated as to be an important factor influencing the receipt of postoperative adjuvant chemotherapy among colon cancer (Jorgensen, Young, Dobbins, & Solomon, 2014). Previous studies suggested that chemotherapy for older patients were less recommend (Jorgensen, Young, & Solomon, 2011). There are other colon cancer studies indicated that there exist a socioeconomic disparity in receipt of chemotherapy. Specifically, patients in the middle of the socioeconomic strata were less likely to receive chemotherapy (Jorgensen et al., 2014). Other factors, such as non-clinical factors including public health

funded status and race may also influence the reception of neo-adjuvant chemotherapy (Chase, Rincon, Deane, Tewari, & Brewster, 2009).

In addition, for the disparity of survival benefit of chemotherapy, the QUASAR trial demonstrated that efficacy of treatment did not significantly differ by tumor site, tumor stage, gender, or age, and the survival benefit with chemotherapy for older patients is less (Group, 2007).

# 3. Methodology

We conducted a retrospective study to evaluate whether the socioeconomic covariates were predictors of receipt of chemotherapy (Aim1), and the survival benefit of adjuvant chemotherapy for stage II colon cancer patients. For patients received adjuvant chemotherapy, this study also investigated whether these covariates were predictors for overall survival (Aim2).

## Data source and study population

This retrospective study was performed on the National Cancer Data Base (NCDB) of stage II colon cancer patients diagnosed between 2004 and 2012. As a nationwide large oncology outcomes database, the National Cancer Data Base, which include approximately 70% of all new patients diagnosed with cancer in the United States (Bilimoria, Stewart, Winchester, & Ko, 2008) provide a good resource for cancer research. “With the joint endeavor of the American College of Surgeons (ACoS) and the American Cancer Society, the National Cancer Data Base provides clinical and demographic data of

patients in 1,500 Commission on Cancer-approved hospitals across the United State (Melvan et al., 2015).” Once patients diagnosed and/or treated at the Commission on Cancer (CoC) approved hospitals, their followed reports will be sent to the NCDB regardless of whether they receive their care at CoC hospitals (Bilimoria et al., 2008). The quality of NCDB was closely verified, and data abstraction is performed by trained cancer registrars who are subject to routine audit by the Commission on Cancer (Asare et al., 2016).

As our study focused on the adjuvant chemotherapy use in stage II colon cancer patients after their surgeries, every case selected into our analysis was treated with surgery after diagnosing. Patients were selected if they underwent a complete resection of colon cancer and shown no evidence of distant metastases.

The Aim1 study population of stage II colon cancer patients were abstracted with other inclusion and exclusion criteria. For example, cases receiving radiation therapy and patients receiving chemotherapy before surgery were eliminated from further analysis. In addition, patients whose histology classified as Adenocarcinoma, squamous, adenosquamous were included into this study, and patients with other histology types were excluded from further analysis. To evaluate whether the socioeconomic factors could predict the overall survival for patients with adjuvant chemotherapy (Aim2), patients would only be included if they received adjuvant chemotherapy after surgery. All selection criteria were listed in Table 1-1 and Table 1-2, which showed the reduction of sample sizes step by step, generated for both Aim 1 population and Aim 2 population.

The primary outcome for Aim 1 was a binary variable, which indicated whether a

patient went through adjuvant chemotherapy after surgery. Using statistical approaches, the association between the socioeconomic factors and the adjuvant chemotherapy usage would be investigated. The outcome for aim 2 analyses was the overall survival time, which defined as the subtraction of the date of last contact or death and the date of their surgical procedure. The survival time was calculated with unit month.

In this study the missing data proportion for most factors was relatively small in regard to the huge sample size. Therefore, no handling methods used for imputing missing data. To analyze the data more efficiently, factors with huge missing data have eliminated from the multivariate analysis.

## Covariates

Patient demographics (age at diagnosis, race, sex, Spanish or Hispanic Origin) and diagnosis and disease/treatment characteristics (Year of diagnosis, Charlson-Deyo comorbidity Score, regional lymph invasion, the presence of positive surgical margins, tumor size, grade, sequence number) were obtained from NCDB data. Socioeconomic factors examined in this study include facility type, facility location, primary payer, median income quartiles 2000, insurance type, percent no high school degree quartiles 2000, urban/rural disparities, and great circle distance.

Race and ethnicity were categorized as white, black and other in race, while using Spanish or Hispanic origin variable to identify persons of Spanish or Hispanic origin. Charlson-Deyo comorbidity Score were grouped as 0, 1 and 2+, while a score of 0 indicates "no comorbid conditions recorded." Facility type in NCDB is assigned to a

category proposed by the Commission on Cancer Accreditation program, which provides a general classification of each reporting facility. At the same time, the NCDB estimates the educational attainment and the annual median income for patients by referencing the zip code of patients residence with US Census 2000 data (Upadhyay, Dahal, Bhatt, Khanal, & Silberstein, 2015).

## Statistical analysis

All statistical analysis in this study were conducted using SAS Version 9.3, and SAS macros developed by Biostatistics and Bioinformatics Shared Resource at Winship Cancer Institute (Nickleach. et al., 2013). The significant level was set at 0.05 for all significance tests. For the model selection procedure, backward selections using $\alpha = 0.1$ removal criteria were conducted for both logistic regression models and Cox regression models.

### Descriptive statistics

We calculated the summary statistics (mean, median, standard deviation) for each numeric variable and frequencies table for categorical variables. Furthermore, number of missing were reported for each variable if applicable.

### Univariate analysis

The univariate association of each interested covariate and receipt of chemotherapy was assessed by chi-square tests for categorical covariates, and ANOVA for numerical covariates (Mikell et al., 2015). The univariate associations between covariates and the overall survival were tested by Cox proportional hazards models and log-rank tests.

The chi-square tests can detect any difference of the mean of responses to the discrete outcome in several independent comparison groups (J. J. Lin, Chang, & Pal, 2015). The null hypothesis is that the distribution of the outcome of the group covariate is independent without association; while the alternative hypothesis is that there exists significant difference in the outcome across the comparison groups. The test statistic is designed for large samples. Specifically, the underline assumption is satisfied as expected frequencies in all response categories in each group are greater than 5. If the expected frequencies were smaller than 5, then Fisher exact tests would be used (Gibbons & Chakraborti, 2011). Since we have a huge sample size, all calculations for categorical covariates were based on chi-square test. The chi-square statistics were calculated as follow:

$$X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(n_{ij} - \hat{\mu}_{ij})^2}{\hat{\mu}_{ij}} \sim X^2_{(r-1)(c-1)}$$

Where $n_{ij}$ denotes observed frequency, $\hat{\mu}_{ij}$denotes expected frequency, r denotes the number of rows in the two-way table and c denotes the number of columns in the two-way table.

Since Analysis of variance (ANOVA) is widely used to analyze the differences among group means (Michalos, 2014), we adopt this approach in this study. ANOVA generalizes the t-test to more than two groups without increasing type I error, and provides a statistical test of detecting the different in means of several groups (Michalos, 2014; Tabachnick, Fidell, & Osterlind, 2001).

**Crude Survival analysis**

First, to crudely compare the two survival functions for patients with and without adjuvant chemotherapy, the unadjusted Log-rank test was conducted.

Furthermore, we performed Log-Rank tests to identify whether there is a difference between the survival functions for the different groups crossing our interested factors. Using Log-rank test to compare two groups of treatments, we artificially used weight equal to 1, and the log-rank statistics appeared a chi-square distribution (Harrell, 2015). Then the formula become as follow:

$$Z = \frac{\sum_1^D d_{i1} - Y_{i1}(\frac{d_i}{Y_i})}{\sqrt{\sum_1^D \frac{Y_{i1}}{Y_i}(1 - \frac{Y_{i1}}{Y_i})(\frac{Y_i - d_i}{Y_{i-1}})d_i}} \qquad Z^2 \sim \chi_1^2$$

**Multivariate analysis:**

Since the first aim of this study was to determine whether the covariates selected from NCDB are predictors of receipt of chemotherapy, the primary outcome is a binary indicator. Therefore, a logistic regression was fitted to evaluate the relationships between receipt of chemotherapy and various covariates. Based on the final selected multivariate logistic model, we would be able to estimate the odds ratio of chemotherapy use across each covariate, with adjusting other potential confounding effect. The logistic regression model was fitted as follows:

$$\text{Logit}\left(\frac{p(Y=1|x)}{p(Y=0|x)}\right) = \beta_0 + \sum_{i=1}^{p} \beta_i X_i$$

Where Y equals to 0 if patients did not receive adjuvant chemotherapy, and 1 if patients received adjuvant chemotherapy.

Our second aim was to investigate whether these covariates were predictors for overall survival for patients received adjuvant chemotherapy. Based on this times to event outcome, a multivariable Cox proportional hazard model was performed. To avoid the bias, patients who did not receive chemotherapy and died within a shorter period after their surgery were considered to be excluded. We calculated the time lag between the surgery and the start of chemotherapy, and eliminated patients died or lose follow up within 3 months (the 75% percentage of the time lag) of their surgery. As the full model may have unimportant parameters, a backward model selection was conducted with $\alpha = 0.1$. The Cox proportional hazard model fitted in this study was as follow:

$$h(t, X) = h_0(t)\exp(\sum_{i=1}^{p} \beta_i X_i)$$

Where $X=(X_1, X_2, \ldots, X_P)$ were the predictor variables.

In addition, to evaluate the benefit of receiving adjuvant chemotherapy adjusted other covariates, a multivariate cox regression model was used to estimate the marginal hazard ratio of receiving adjuvant chemotherapy verses not receiving chemo treatment.

**Propensity Score matching**

To reduce the selection bias and potential confounding effect, a propensity score matching approach was adopted (Rosenbaum & Rubin, 1983). A logistic regression model predicting adjuvant chemotherapy alone vs. non-Chemo was carried out to

estimate the propensity score by all baseline covariates of interest. Patients from each study cohort were matched to each other at ratio of 1:1 based on the propensity score using a greedy 5-1 digit match algorithm (Parsons. & Group., 2001). After matching, the balance of covariate between two cohorts was evaluated by the standardized differences and a value of < 0.1 was considered as negligible imbalance (Austin, Grootendorst, & Anderson, 2007). The effects were estimated in the matched sample by a Cox hazard model with a robust variance estimator for overall survival (D. Y. Lin & Wei, 1989). Finally, a repeated univariate survival analysis for the efficacy of chemotherapy was conducted based on matched dataset.

# 4. Results

## 4.1 Aim 1

### Demographics

The NCDB colon cancer database included 712,172 patients. Based on our selection criteria, there were 103,660 patients met all selection criteria, and were included into the statistical analysis for Aim 1. The selection diagrams for Aim 1 was shown in the appendix as Table 1-1. The demographics of the aim 1 and aim 2 study cohorts were provided. Through the descriptive statistics, we known that the majority patients in the study population for aim 1 were approximately 70 years old, white, with a single malignant primary tumor and lived in metropolitan counties. Besides, most of the patients did not show a positive in lymph vascular invasion, microsatellite instability, and surgical margins. The study cohort also distributed equally through sex, facility location, and

socioeconomic factors such as income and education status. For the insurance payer, 30% patients were covered by private insurance, while the percentage of Medicaid and Medicare were 3.5% and 61.1%, respectively.

**Unadjusted analysis of receipt chemotherapy**

We first performed a univariate analysis based on the aim 1 cohort to evaluate the association of variables with the receipt of adjuvant chemotherapy after surgery. The result was shown in Table 3-3. As the result shown, some demographic variables associated with the increasing receipt of chemotherapy. For example, patients who were black, male, Spanish or Hispanic, and relatively younger age group shown a higher probability of receipt the adjuvant chemotherapy after surgery (all p-value <0.001). Clinical variables were also shown associated with increasing postoperative usage of adjuvant chemotherapy. Without adjusting other variables, patients with large tumor size, positive surgical margins, positive lymph vascular invasion, AJCC pathologic stage in 2B or 2C, sequence number in 0, and Charlson-Deyo comorbidity score equal to 0 tend to receive adjuvant chemotherapy.

The crude association of socioeconomic variables with receipt of chemotherapy was also studied. Our analysis showed that patients with small great circle distance, treated in northeast, lived in urban area, and low education attainment area tend to receive adjuvant chemotherapy after colon cancer resection without adjusted other effects.

**Crude survival analysis**

In crude survival analysis of Aim 1 has shown in Table 3-1. Based on the Table 3-1,

we found there exist a significant survival benefit for patients with chemotherapy.

Also, patients who were young of age, non-black, female, Spanish or Hispanic, had significant overall survival advantage. There existed association between socioeconomic factors and the overall survival. For example, patients lived in area with high income and education level had a relative better overall survival. Additionally, compared with patient without insurance, we found private insurance had a high survival rate, while patients with Medicare had a worst survival rate. Besides, patients treated in "community cancer program or other" shown lower survival rate than treated in "comprehensive community cancer program" or "integrated network cancer program." The results also shown that patients treated in academic program had the highest survival rate. In addition，clinical variables including sequence number equal to 0 and Charlson-Deyo comorbidity Score equal to 0, negative surgical margins and negative lymphovascular invasion were associated with a better overall survival. We observed that the survival rate continue increasing slightly through diagnosis year.

**Multivariate analysis of receipt chemotherapy**

This study performed a multivariable analysis to identify variables that might consider as predictor of receipt adjuvant chemotherapy. The results in Table 4-1 demonstrated that younger age (OR=0.92 with 95% CI: 0.92 to 0.93), white, Spanish or Hispanic were associated with greater odds of receiving adjuvant chemotherapy, compared to the counterpart. In addition, patients with large tumor size (OR=1.02 with 95%CI: 0.95 to 0.98), positive surgical margins (OR=2.63 with 95% CI: 2.42 to 2.86), sequence number equal to 1, and Charlson-Deyo comorbidity score equal to 0 were

associated with a higher odds of receiving adjuvant chemotherapy. Additionally, socioeconomic variables played an essential role for patients of whether receive adjuvant chemotherapy. Our data shown that patients with small great circle distance (OR=0.96 with 95% CI: 0.95 to 0.98), treated with facility in northeast, and lived in a low-income area tend to receive adjuvant chemotherapy after colon cancer resection. Compared insurance with Medicaid, private insurance or not insured, patients insured with Medicare had a greater odds of receiving adjuvant chemotherapy. Besides, patients treated in community cancer program or other shown higher odds than treated in comprehensive community cancer program, integrated network cancer program, or academic program. We also found that the general usage of adjuvant chemotherapy decrease in the past 8 years.

**Multivariate survival analysis**

For aim 1 analysis, the landmark multivariate Cox regression model eliminated patients that died within 2 months after surgery. The results indicated that with adjusting for all other confounding available in our dataset, there exist a significant benefit of receiving postoperative adjuvant chemotherapy for patients with stage II colon cancer of overall survival (HR: 0.82 with 95% CI: 0.79 to 0.86). In addition, our data confirmed that patients in young age, non-black, female, and Spanish or Hispanic have a trend of better survival. Clinical factors including sequence number equal to 0, small Charlson-Deyo score, negative surgical margins associated with a better survival. For socioeconomic factors, the analysis demonstrated that patients lived in high-income and high-educated areas have a better survival. Additionally, patients insured with private insurance (HR: 0.86 with 95% CI: 0.77 to 0.96) and Medicare (HR: 0.96 with 95% CI:

0.86 to 1.07) shown a better survival than patients without insurance (HR: 1.32 with 95% CI: 1.14 to 1.53). Additional information was shown in Table 4-2.

## 4.2 Aim 2

### Demographics

Among the Aim 1 study cohort, 19,497 were treated with adjuvant chemotherapy (Aim 2). The selection diagrams for Aim 2 was shown in the appendix as Table 1-2. In our aim 2 cohort, the majority patients were approximately 60 years old, white, insured with private insurance (51%), living in the metropolitan area, negative surgical margins, with sequence number equal to 0 and Charlson-Deyo score equal to 0. Additional demographics information was enclosed in the Table 2-1 and Table 2-2, respectively.

### Crude survival analysis

In crude survival analysis of Aim 2 has shown in Table 3-2. This study confirmed the univariate association of variables with overall survival for patients received adjuvant chemotherapy after surgery (Aim2). As the results shown in Table 3-4, the benefit of chemotherapy was relative higher for patients in young age, other race, female, Spanish or Hispanic, Clinical factors including sequence number equal to 0, Charlson-Deyo comorbidity Score equal to 0, and negative surgical margins expressed a better overall survival after chemotherapy. For socioeconomic factors, we found that patients insured with private insurance or other government insurance, participated in academic and research program, treated in west of the United States tend to have a better survival with chemotherapy. Also, patients lived in high-income community and high-education

attainment area had a greater advantage after chemotherapy.

### Multivariate survival analysis

The Cox regression model for Aim 2 continued shown an improvement in overall survival with young age, non-black, female, negative surgical margins, sequence number equal to 0, and Charlson-Deyo comorbidity Score equal to 0. For socioeconomic factors, our data demonstrated a significant association of small great circle distance, lived in high-income community with better overall survival. There was also a trend toward better overall survival for patients insured with private insurance, compared to all other payer. We also found that patients with Medicaid (HR: 1.20 with 95% CI: 0.89 to 1.63) had a worse survival compared to other payer. Furthermore, patients with academic or research program shown the best survival compared to the counterpart of facility types. Additional results shown in Table 4-3.

## 4.3 Other findings

### Propensity score match analysis

Based on the original research dataset, the crude extended model confirmed the benefit of adjuvant chemotherapy in stage II colon cancer (HR: 0.40 with 95% CI: 0.38 to 0.41). To obtain a more powerful result, with control all other baseline effect, a propensity score matching was conducted. (The distributions of propensity score were showed in Figure 1.)

Propensity score matching identified 16548 matched pairs, for a total of 33096 patients. According to the value of standardized difference (Table 5-1), we concluded that

no differences were noted in all potential confounding factors, indicating that the matching procedure worked well. Table 5-2 showed the distributions for survival benefit of the chemo and non-chemo groups after matching. After PS matching, patients with adjuvant chemotherapy after surgery shown a significantly benefit in overall survival (HR: 0.72 with 95% CI: 0.69 to 0.76, $P < 0.001$).

**Stratify analysis**

Potential interaction effects based on the matched dataset were also assessed. As the Table 6, we found that the benefit of chemotherapy in survival is relatively high for patients in non-black, insured with Medicaid or Medicare, lived in the high-income and high-educated area, and positive surgical margins.

# 5. Discussion

## Summary of study & Conclusion

This study is the largest nationwide analysis of postoperative adjuvant chemotherapy in stage II colon cancer. Although chemotherapy is one of the standard adjuvant care for patients with colon cancer after their resection, there exist a disparity in receipt of chemotherapy. It is unclear whether there exist some predictors that associated with the usage of postoperative chemotherapy. Therefore, for further understanding disparities in receipt of adjuvant chemotherapy (Aim1), our study based on NCDB database demonstrated that, similar to previous studies, demographic factors, such as younger age, non-black, female sex, and Spanish or Hispanic were related a trend of higher

chemotherapy usage. More importantly, this study confirmed that socioeconomic factors including treated in non-west facility, living in high-income neighborhoods and communities with a higher proportion of high school graduates were significantly increase the chance of receipt chemotherapy. Additionally, insured patients, especially insured with private companies have a greater percentage in receiving chemotherapy compared to uninsured patients.

In addition, survival analyses were conduct to study whether there is some predictor that indicating a better overall survival. Our results indicated that demographic factors including younger age, female sex were associated with higher survival rate. Clinical factors, such as small tumor size, negative surgical margins, lower sequence number and Charlson-Deyo score were associated with better survival. Besides, better socioeconomic situation, such as living in high-income neighborhoods with a higher proportion of high school graduates were significantly increasing the overall survival.

Although it is generally recommended that patients with stage III patients to receive postoperative chemotherapy, the survival benefit in stage II colon cancer is no consensus (Dienstmann et al., 2015). There is no specific study that provides a clear survival benefit for patients who have stage II colon cancer based the national and multicenter database(McKenzie et al., 2011). After adjusting baseline effect by propensity score matching, postoperative chemotherapy was still associated with improved overall survival in stage II colon cancer. Our results suggested that surgery followed by chemotherapy is acceptable and provide additional benefit for patience with stage II colon cancer. Although as the Quasar trial shown, the absolute benefits are not extremely

large (Group, 2007). Additionally, this study assessed the survival benefit of chemotherapy in stage II colon cancer for patients with different demographic, socioeconomic and clinical factors.

## Limitations

The National Cancer Data Base confers a large sample size, which allow us to do the propensity score matching and break down into subgroup. However, the NCDB is limited by its inherent retrospective nature, the potential coding error(Mikell et al., 2015). Beside, there also are other drawbacks in this study. As the NCDB data estimates the educational attainment and the annual median income for patients by referencing the zip code of patients residence with US Census 2000 data (Upadhyay et al., 2015), we are not able to obtain individual-level income and education attainments, which made our result become vague and less powerful. Also, due to high missing proportion in key pathological factors including the lymph vascular invasion, microsatellite instability and perineural invasion, this study could not adjust their effect in the multivariate models, which may result in less-meaningful or misleading conclusion. Finally, the pattern and cause for patients missing in these key pathological factors were unclear in this analysis.

## Future research

In this study, we attempted to determine whether adjuvant chemotherapy significantly lowered the risk of death. The result of overall survival could be misleading, because patients who died quickly had less time available to get chemotherapy. This kind of lead-time bias could be handled by extended survival model with a time dependent

covariate or landmark model. Due to huge sample size, this study adopted the landmark survival model for multivariate analysis instead of the extended survival model. Therefore, to obtain a more accurate result of the survival benefit of chemotherapy, future research could use other advanced survival models to optimize the approach of multivariate survival analysis.

Also, due to huge missing in the National Cancer Data Base, important pathological factors including Perineural invasion, Lymphovascular invasion, Microsatellite instability could not be added into the multivariate models. As the previous studies showed, these factors are highly related to the overall survival and decision making of whether receive a chemotherapy for stage II colon cancer patients (Artac et al., 2014; C. C. Lin et al., 2014; Yun et al., 2014), future research could focus on other nationwide dataset with complete pathological data, to evaluate the exact association between socioeconomic factors and the receipt of chemotherapy while adjusting these high-risk pathological factors.

The national cancer database is a comprehensive database, which includes valuable information of patients and their treatment, but the number of exploratory factors is not enough to build a predicting model. It would be very impressive and useful if a new index or predicted model of receipt of chemotherapy in stage II colon cancer could be created.

# 6. Reference

André, T., Boni, C., Mounedji-Boudiaf, L., Navarro, M., Tabernero, J., Hickish, T., . . . Bridgewater, J. (2004). Oxaliplatin, fluorouracil, and leucovorin as adjuvant treatment for colon cancer. *New England Journal of Medicine, 350*(23),

2343-2351.

Artac, M., Turhal, N. S., Kocer, M., Karabulut, B., Bozcuk, H., Yalcin, S., . . . Uygun, K. (2014). Do high-risk features support the use of adjuvant chemotherapy in stage II colon cancer? A Turkish Oncology Group study. *Tumori, 100*(2), 143-148. doi: 10.1700/1491.16397

Asare, E. A., Liu, L., Hess, K. R., Gordon, E. J., Paruch, J. L., Palis, B., . . . Bilimoria, K. Y. (2016). Development of a model to predict breast cancer survival using data from the National Cancer Data Base. *Surgery, 159*(2), 495-502. doi: 10.1016/j.surg.2015.08.006

Austin, P. C., Grootendorst, P., & Anderson, G. M. (2007). A comparison of the ability of different propensity score models to balance measured variables between treated and untreated subjects: a Monte Carlo study. *Stat Med, 26*(4), 734-753. doi: 10.1002/sim.2580

Benson, A. B., Schrag, D., Somerfield, M. R., Cohen, A. M., Figueredo, A. T., Flynn, P. J., . . . Van Cutsem, E. (2004). American Society of Clinical Oncology recommendations on adjuvant chemotherapy for stage II colon cancer. *Journal of clinical oncology, 22*(16), 3408-3419.

Bilimoria, K. Y., Stewart, A. K., Winchester, D. P., & Ko, C. Y. (2008). The National Cancer Data Base: a powerful initiative to improve cancer care in the United States. *Ann Surg Oncol, 15*(3), 683-690. doi: 10.1245/s10434-007-9747-3

Chase, D. M., Rincon, A., Deane, M., Tewari, K. S., & Brewster, W. R. (2009). Socioeconomic factors may contribute to neoadjuvant chemotherapy use in metastatic epithelial ovarian carcinoma. *Gynecol Oncol, 115*(3), 339-342. doi: 10.1016/j.ygyno.2009.08.008

DeSantis, C. E., Lin, C. C., Mariotto, A. B., Siegel, R. L., Stein, K. D., Kramer, J. L., . . . Jemal, A. (2014). Cancer treatment and survivorship statistics, 2014. *CA: a cancer journal for clinicians, 64*(4), 252-271.

Dienstmann, R., Salazar, R., & Tabernero, J. (2015). Personalizing colon cancer adjuvant therapy: selecting optimal treatments for individual patients. *J Clin Oncol, 33*(16), 1787-1796. doi: 10.1200/jco.2014.60.0213

Gibbons, J. D., & Chakraborti, S. (2011). *Nonparametric statistical inference*: Springer.

Gill, S., Loprinzi, C. L., Sargent, D. J., Thome, S. D., Alberts, S. R., Haller, D. G., . . . Goldberg, R. M. (2004). Pooled analysis of fluorouracil-based adjuvant therapy for stage II and III colon cancer: who benefits and by how much? *J Clin Oncol, 22*(10), 1797-1806. doi: 10.1200/jco.2004.09.059

Gill, S., Loprinzi, C. L., Sargent, D. J., Thomé, S. D., Alberts, S. R., Haller, D. G., . . . Seitz, J. F. (2004). Pooled analysis of fluorouracil-based adjuvant therapy for stage II and III colon cancer: who benefits and by how much? *Journal of clinical oncology, 22*(10), 1797-1806.

Group, Q. C. (2007). Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study. *The Lancet, 370*(9604), 2020-2029.

Haggar, F. A., & Boushey, R. P. (2009). Colorectal Cancer Epidemiology: Incidence, Mortality, Survival, and Risk Factors. *Clin Colon Rectal Surg, 22*(4), 191-197.

Harrell, F. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*: Springer.

Jorgensen, M. L., Young, J. M., Dobbins, T. A., & Solomon, M. J. (2014). Does patient age still affect receipt of adjuvant therapy for colorectal cancer in New South Wales, Australia? *J Geriatr Oncol, 5*(3), 323-330. doi: 10.1016/j.jgo.2014.02.007

Jorgensen, M. L., Young, J. M., & Solomon, M. J. (2011). Older patients and adjuvant therapy for colorectal cancer: surgeon knowledge, opinions, and practice. *Dis Colon Rectum, 54*(3), 335-341. doi: 10.1007/DCR.0b013e3181ff43d6

Kamangar, F., Dores, G. M., & Anderson, W. F. (2006). Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *Journal of clinical oncology, 24*(14), 2137-2150.

Kerr, D. (1989). 5-Fluorouracil and folinic acid: interesting biochemistry or effective treatment? *British journal of cancer, 60*(6), 807.

Lin, C. C., Lin, J. K., Lin, T. C., Chen, W. S., Yang, S. H., Wang, H. S., . . . Chang, S. C. (2014). The prognostic role of microsatellite instability, codon-specific KRAS, and BRAF mutations in colon cancer. *J Surg Oncol, 110*(4), 451-457. doi: 10.1002/jso.23675

Lin, D. Y., & Wei, L. J. (1989). The Robust Inference for the Cox Proportional Hazards Model. *Journal of the American Statistical Association, 84*, 1074--1078. doi: 10.2307/2290085

Lin, J. J., Chang, C. H., & Pal, N. (2015). A revisit to contingency table and tests of independence: bootstrap is preferred to Chi-square approximations as well as Fisher's exact test. *J Biopharm Stat, 25*(3), 438-458. doi: 10.1080/10543406.2014.920851

Mamounas, E., Wieand, S., Wolmark, N., Bear, H. D., Atkins, J. N., Song, K., . . . Rockette, H. (1999). Comparative efficacy of adjuvant chemotherapy in patients with Dukes' B versus Dukes' C colon cancer: results from four National Surgical Adjuvant Breast and Bowel Project adjuvant studies (C-01, C-02, C-03, and C-04). *Journal of clinical oncology, 17*(5), 1349-1349.

Mamounas, E. W., S. Wolmark, N., Bear, H. D., Atkins, J. N., & Song, K. (1999). Comparative efficacy of adjuvant fluorouracil and folinic acid in B2 colon cancer. International Multicentre Pooled Analysis of B2 Colon Cancer Trials (IMPACT B2) Investigators. *J Clin Oncol, 17*(5), 1356-1363.

McKenzie, S., Nelson, R., Mailey, B., Lee, W., Chung, V., Shibata, S., . . . Kim, J. (2011). Adjuvant chemotherapy improves survival in patients with American Joint Committee on Cancer stage II colon cancer. *Cancer, 117*(24), 5493-5499. doi: 10.1002/cncr.26245

Melvan, J. N., Sancheti, M. S., Gillespie, T., Nickleach, D. C., Liu, Y., Higgins, K., . . . Fernandez, F. G. (2015). Nonclinical Factors Associated with 30-Day Mortality after Lung Cancer Resection: An Analysis of 215,000 Patients Using the National Cancer Data Base. *J Am Coll Surg, 221*(2), 550-563. doi: 10.1016/j.jamcollsurg.2015.03.056

Michalos, A. C. (2014). *Encyclopedia of Quality of Life and Well-Being Research*.

Mikell, J. L., Gillespie, T. W., Hall, W. A., Nickleach, D. C., Liu, Y., Lipscomb, J., . . . Higgins, K. A. (2015). Postoperative radiotherapy is associated with better survival in

non-small cell lung cancer with involved N2 lymph nodes: results of an analysis of the National Cancer Data Base. *J Thorac Oncol, 10*(3), 462-471. doi: 10.1097/jto.0000000000000411

Nickleach., D., Liu., Y., Shrewsberry., A., Ogan., K., Kim;., S., & Wang, Z. (2013). SAS® Macros to Conduct Common Biostatistical Analyses and Generate Reports. SESUG 2013: The Proceeding of the SouthEast SAS User Group. .

Organization, W. H. (2002). *Cancer Incidence in Five Continents*. The World Health Organization

Lyon, France.

Parsons., L. S., & Group., O. R. (2001). Reducing bias in a propensity score matched-pair sample using greedy matching techniques: Seattle, WA.

Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*(1), 41-55.

Sharma, S., & O'Keefe, S. J. D. (2007). Environmental influences on the high mortality from colorectal cancer in African Americans. *Postgrad Med J, 83*(983), 583-589.

Tabachnick, B. G., Fidell, L. S., & Osterlind, S. J. (2001). Using multivariate statistics.

Upadhyay, S., Dahal, S., Bhatt, V. R., Khanal, N., & Silberstein, P. T. (2015). Chemotherapy use in stage III colon cancer: a National Cancer Database analysis. *Ther Adv Med Oncol, 7*(5), 244-251. doi: 10.1177/1758834015587867

Yun, J. A., Kim, H. C., Kim, S. H., Cho, Y. B., Yun, S. H., Lee, W. Y., & Chun, H. K. (2014). Prognostic significance of perineural invasion in stage IIA colon cancer. *ANZ J Surg*. doi: 10.1111/ans.12810

Zhang, Y., Gao, H., Bu, Y., Fan, X., & Jia, J. (2013). Factors associated with receipt of adjuvant chemotherapy among married women with breast cancer. *World J Surg Oncol, 11*, 286. doi: 10.1186/1477-7819-11-286

# 7. Appendices

Table 1-1. Selection Diagram for Aim 1

| Selection and Exclusion Criteria | Sample Size | Excluded |
|---|---|---|
| NCDB Colon PUF Cancer Cases | 712172 | - |
| Include patients with surgery | 590587 | 121585 |
| Exclude patients with radiation | 579977 | 10610 |
| Exclude patients had Chemotherapy before Surgery | 573228 | 6749 |
| Include patients with stage 2 colon cancer | 159657 | 413571 |
| Exclude patients whose colon cancer was in other stage | 159160 | 497 |
| Exclude if tnm_path_n in ('1','1A','1B','1C','2','2A','2B') | 158509 | 651 |
| Exclude if Regional Lymph Nodes was Positive | 158178 | 331 |
| Include if BEHAVIOR was invasive | 158164 | 14 |
| Exclude patients if Chemotherapy situation was unknown | 150188 | 7976 |
| Exclude patients with distant metastasis | 149658 | 530 |
| Include Histology: adeno, squamous, adenosquamous | 145856 | 3802 |
| Exclude if patient has palliative care | 145856 | 0 |
| Exclude if Outcome (survival) is missing | 128188 | 17668 |
| Include if patients had a single malignant primary tumor | 103660 | 24528 |

Table 1-2. Selection/Exclusion Diagram for Aim 2

| **Selection and Exclusion Criteria** | **Sample Size** | **Excluded** |
|---|---|---|
| NCDB Colon PUF Cancer Cases | 712172 | - |
| Include patients with surgery | 590587 | 121585 |
| Exclude patients with radiation | 579977 | 10610 |
| Exclude patients had Chemotherapy before Surgery | 573228 | 6749 |
| Include patients with stage 2 colon cancer | 159657 | 413571 |
| Exclude patients whose colon cancer was in other stage | 159160 | 497 |
| Exclude if tnm_path_n in ('1','1A','1B','1C','2','2A','2B') | 158509 | 651 |
| Exclude if Regional Lymph Nodes was Positive | 158178 | 331 |
| Include if BEHAVIOR was invasive | 158164 | 14 |
| Exclude patients if Chemotherapy situation was unknown | 150188 | 7976 |
| Exclude patients with distant metastasis | 149658 | 530 |
| Include Histology: adeno, squamous, adenosquamous | 145856 | 3802 |
| Exclude if patient has palliative care | 145856 | 0 |
| Exclude if Outcome(survival) is missing | 128188 | 17668 |
| Include if patients had a single malignant primary tumor | 103660 | 24528 |
| Exclude patients did not have Chemotherapy | 19497 | 84163 |

Table 2-1. Descriptive Statistics For Aim 1

| Variable | Level | N (%) = 103140 |
|---|---|---|
| Race | White | 87610 (84.9) |
| | Black | 11433 (11.1) |
| | other | 4097 (4.0) |
| Sex | Male | 48294 (46.8) |
| | Female | 54846 (53.2) |
| Spanish Hispanic Origin | Non-Spanish; non-Hispanic | 90530 (87.8) |
| | Spanish or Hispanic | 4790 (4.6) |
| | Unknown | 7820 (7.6) |
| Year of Diagnosis | 2004 | 11927 (11.6) |
| | 2005 | 11786 (11.4) |
| | 2006 | 11858 (11.5) |
| | 2007 | 11513 (11.2) |
| | 2008 | 11335 (11.0) |
| | 2009 | 11126 (10.8) |
| | 2010 | 10864 (10.5) |
| | 2011 | 11293 (10.9) |
| | 2012 | 11438 (11.1) |
| Primary Payer | Not Insured | 3385 (3.3) |
| | Private Insurance | 30952 (30.0) |
| | Medicaid | 3659 (3.5) |
| | Medicare | 62980 (61.1) |
| | Other Government | 657 (0.6) |
| | Insurance Status Unknown | 1507 (1.5) |
| Facility Type | Community Cancer Program/Other | 15834 (15.7) |
| | Comprehensive Community Cancer Program | 53165 (52.6) |
| | Academic/Research Program | 24460 (24.2) |
| | Integrated Network Cancer Program | 7674 (7.6) |
| | Missing | 2007 |
| Facility Location | Northeast | 20135 (19.9) |
| | South | 37921 (37.5) |
| | Midwest | 28346 (28.0) |
| | West | 14731 (14.6) |
| | Missing | 2007 |

| Variable | Level | N (%) = 103140 |
|---|---|---|
| Median Income Quartiles 2000 | Not Available | 3471 |
| | < $30,000 | 13884 (13.9) |
| | $30,000 - $35,999 | 18354 (18.4) |
| | $36,000 - $45,999 | 28556 (28.7) |
| | $46,000 + | 38875 (39.0) |
| Percent No High School Degree Quartiles 2000 | Not Available | 3478 |
| | >=29% | 16889 (16.9) |
| | 20-28.9% | 23633 (23.7) |
| | 14-19.9% | 24071 (24.2) |
| | < 14% | 35069 (35.2) |
| Urban/Rural 2003 | 1-Metro | 83154 (83.3) |
| | 2-Urban | 14514 (14.5) |
| | 3-Rural | 2179 (2.2) |
| | Missing | 3293 |
| Sequence Number | 0 | 92297 (89.5) |
| | 1 | 10843 (10.5) |
| Charlson-Deyo Score | 0 | 69130 (67.0) |
| | 1 | 24313 (23.6) |
| | 2+ | 9697 (9.4) |
| Grade | Well differentiated, differentiated, NOS | 9843 (9.5) |
| | Moderately differentiated, moderately well differentiated, intermediate differentiation | 73700 (71.5) |
| | Poorly differentiated | 15724 (15.2) |
| | Undifferentiated, anaplastic | 1701 (1.6) |
| | Cell type not determined, not stated or not applicable, unknown primaries, high grade dysplasia | 2172 (2.1) |
| Surgical Margins Status at any CoC Facility | no | 98742 (95.7) |
| | yes | 3610 (3.5) |
| | unknown | 788 (0.8) |
| Chemotherapy at any CoC Facility | No | 83724 (81.2) |
| | Yes | 19416 (18.8) |

| Variable | Level | N (%) = 103140 |
|---|---|---|
| Primary Site | C180 Cecum | 22853 (22.2) |
| | C181 Appendix | 1684 (1.6) |
| | C182 Ascending colon | 24287 (23.5) |
| | C183 Hepatic flexure of colon | 5986 (5.8) |
| | C184 Transverse colon | 11258 (10.9) |
| | C185 Splenic flexure of colon | 4050 (3.9) |
| | C186 Descending colon | 6334 (6.1) |
| | C187 Sigmoid colon | 23879 (23.2) |
| | C188 Overlapping lesion of colon | 1342 (1.3) |
| | C189 Colon, NOS | 1467 (1.4) |
| Surgical Approach at this Facility | No surgical procedure of primary site | 1479 (4.4) |
| | Robotic assisted | 548 (1.6) |
| | Robotic converted to open | 63 (0.2) |
| | Laparoscopic | 10747 (32.0) |
| | Laparoscopic converted to open | 1920 (5.7) |
| | Open or approach unspecified | 18838 (56.1) |
| | Missing | 69545 |
| AJCC Pathologic Stage Group | 2 | 2083 (2.0) |
| | 2A | 88711 (86.0) |
| | 2B | 10526 (10.2) |
| | 2C | 1820 (1.8) |
| Lymph Vascular Invasion | Missing | 69545 |
| | Not present | 26742 (79.6) |
| | present | 4365 (13.0) |
| | Not applicable or Unknown | 2488 (7.4) |
| 30 Day Mortality | No | 97705 (94.7) |
| | Yes | 3884 (3.8) |
| | Missing | 1551 (1.5) |
| 90 Day Mortality | No | 94912 (92.0) |
| | Yes | 6077 (5.9) |
| | Missing | 2151 (2.1) |

| Variable | Level | N (%) = 103140 |
|---|---|---|
| Readmission Within 30 Days of Surgical Discharge | No | 95164 (92.3) |
| | Yes | 5589 (5.4) |
| | Missing | 2387 (2.3) |
| | No | 95164 (92.3) |
| | Yes | 5589 (5.4) |
| | Missing | 2387 (2.3) |
| CS SSF4 - Tumor Deposits | No | 30803 (29.9) |
| | Yes | 431 (0.4) |
| | Missing | 71906 (69.7) |
| CS SSF7 - Microsatellite Instability (MSI) | Missing | 94942 |
| | Negative | 5651 (68.9) |
| | positive | 2547 (31.1) |
| CS SSF8 - Perineural Invasion | Missisng | 72429 |
| | Not present | 28696 (93.4) |
| | Present | 2015 (6.6) |
| Age at Diagnosis | Mean | 70.00 |
| | Median | 72.00 |
| | Std Dev | 13.38 |
| | Missing | 0.00 |
| Tumor Size | Mean | 5.38 |
| | Median | 4.90 |
| | Std Dev | 4.28 |
| | Missing | 3427.00 |
| CS SSF6 - Circumferential Resection Margin (CRM) | Mean | 897.85 |
| | Median | 988.00 |
| | Std Dev | 269.36 |
| | Missing | 0.00 |
| Great Circle Distance | Mean | 0.40 |
| | Median | 0.14 |
| | Std Dev | 1.61 |
| | Missing | 1453.00 |

Table 2-2. Descriptive Statistics For Aim 2

| Variable | Level | N (%) = 19416 |
|---|---|---|
| Race | White | 16070 (82.8) |
| | Black | 2498 (12.9) |
| | other | 848 (4.4) |
| Sex | Male | 9707 (50.0) |
| | Female | 9709 (50.0) |
| Spanish Hispanic Origin | Non-Spanish; non-Hispanic | 16540 (85.2) |
| | Spanish or Hispanic | 1243 (6.4) |
| | Unknown | 1633 (8.4) |
| Year of Diagnosis | 2004 | 2428 (12.5) |
| | 2005 | 2374 (12.2) |
| | 2006 | 2505 (12.9) |
| | 2007 | 2324 (12.0) |
| | 2008 | 2108 (10.9) |
| | 2009 | 2034 (10.5) |
| | 2010 | 1810 (9.3) |
| | 2011 | 1925 (9.9) |
| | 2012 | 1908 (9.8) |
| Primary Payer | Not Insured | 1130 (5.8) |
| | Private Insurance | 9894 (51.0) |
| | Medicaid | 1115 (5.7) |
| | Medicare | 6812 (35.1) |
| | Other Government | 163 (0.8) |
| | Insurance Status Unknown | 302 (1.6) |
| Facility Type | Community Cancer Program/Other | 3019 (16.5) |
| | Comprehensive Community Cancer Program | 9310 (50.8) |
| | Academic/Research Program | 4642 (25.3) |
| | Integrated Network Cancer Program | 1366 (7.4) |
| | Missing | 1079 |
| Facility Location | Northeast | 3805 (20.8) |
| | South | 7008 (38.2) |
| | Midwest | 5196 (28.3) |
| | West | 2328 (12.7) |
| | Missing | 1079 |

| Variable | Level | N (%) = 19416 |
|---|---|---|
| Median Income Quartiles 2000 | Not Available | 633 |
| | < $30,000 | 2774 (14.8) |
| | $30,000 - $35,999 | 3477 (18.5) |
| | $36,000 - $45,999 | 5206 (27.7) |
| | $46,000 + | 7326 (39.0) |
| Percent No High School Degree Quartiles 2000 | Not Available | 635 |
| | >=29% | 3395 (18.1) |
| | 20-28.9% | 4621 (24.6) |
| | 14-19.9% | 4409 (23.5) |
| | < 14% | 6356 (33.8) |
| Urban/Rural 2003 | 1-Metro | 15614 (82.9) |
| | 2-Urban | 2809 (14.9) |
| | 3-Rural | 414 (2.2) |
| | Missing | 579 |
| Sequence Number | 0 | 17465 (90.0) |
| | 1 | 1951 (10.0) |
| Charlson-Deyo Score | 0 | 14975 (77.1) |
| | 1 | 3599 (18.5) |
| | 2+ | 842 (4.3) |
| Grade | Well differentiated, differentiated, NOS | 1744 (9.0) |
| | Moderately differentiated, moderately well differentiated, intermediate differentiation | 13203 (68.0) |
| | Poorly differentiated | 3546 (18.3) |
| | Undifferentiated, anaplastic | 369 (1.9) |
| | Cell type not determined, not stated or not applicable, unknown primaries, high grade dysplasia | 554 (2.9) |
| Surgical Margins Status at any CoC Facility | no | 17978 (92.6) |
| | yes | 1197 (6.2) |
| | unknown | 241 (1.2) |

| Variable | Level | N (%) = 19416 |
|---|---|---|
| Primary Site | C180 Cecum | 3849 (19.8) |
| | C181 Appendix | 634 (3.3) |
| | C182 Ascending colon | 3540 (18.2) |
| | C183 Hepatic flexure of colon | 957 (4.9) |
| | C184 Transverse colon | 2068 (10.7) |
| | C185 Splenic flexure of colon | 886 (4.6) |
| | C186 Descending colon | 1475 (7.6) |
| | C187 Sigmoid colon | 5403 (27.8) |
| | C188 Overlapping lesion of colon | 266 (1.4) |
| | C189 Colon, NOS | 338 (1.7) |
| Surgical Approach at this Facility | No surgical procedure of primary site | 739 (13.1) |
| | Robotic assisted | 69 (1.2) |
| | Robotic converted to open | 13 (0.2) |
| | Laparoscopic | 1426 (25.3) |
| | Laparoscopic converted to open | 322 (5.7) |
| | Open or approach unspecified | 3074 (54.5) |
| | Missing | 13773 |
| AJCC Pathologic Stage Group | 2 | 378 (1.9) |
| | 2A | 14256 (73.4) |
| | 2B | 3955 (20.4) |
| | 2C | 827 (4.3) |
| Lymph Vascular Invasion | Missing | 13773 |
| | Not present | 3976 (70.5) |
| | present | 1120 (19.8) |
| | Not applicable or Unknown | 547 (9.7) |
| 30 Day Mortality | No | 19342 (99.6) |
| | Yes | 5 (0.0) |
| | Missing | 69 (0.4) |
| 90 Day Mortality | No | 19195 (98.9) |
| | Yes | 81 (0.4) |
| | Missing | 140 (0.7) |
| Readmission Within 30 Days of Surgical Discharge | No | 17874 (92.1) |
| | Yes | 875 (4.5) |
| | Missing | 667 (3.4) |

| Variable | Level | N (%) = 19416 |
|---|---|---|
| CS SSF4 - Tumor Deposits | No | 5019 (25.8) |
| | Yes | 121 (0.6) |
| | Missing | 14276 (73.5) |
| CS SSF7 - Microsatellite Instability (MSI) | Missing | 17668 |
| | Negative | 1340 (76.7) |
| | positive | 408 (23.3) |
| CS SSF8 - Perineural Invasion | Missisng | 14477 |
| | Not present | 4403 (89.1) |
| | Present | 536 (10.9) |
| Age at Diagnosis | Mean | 59.75 |
| | Median | 60.00 |
| | Std Dev | 12.25 |
| | Missing | 0.00 |
| Tumor Size | Mean | 5.79 |
| | Median | 5.00 |
| | Std Dev | 4.34 |
| | Missing | 964.00 |
| CS SSF6 - Circumferential Resection Margin (CRM) | Mean | 909.34 |
| | Median | 988.00 |
| | Std Dev | 257.65 |
| | Missing | 0.00 |
| Great Circle Distance | Mean | 0.37 |
| | Median | 0.15 |
| | Std Dev | 1.36 |
| | Missing | 230.00 |

Table 3-1. Univariate Survival Analysis for Receipt of Chemotherapy

| | | | Chemotherapy at any CoC Facility | | |
|---|---|---|---|---|---|
| **Covariate** | **Statistics** | **Level** | **No N=83724** | **Yes N=19416** | **Parametric P-value*** |
| Race | N (Row %) | White | 71540 (81.66) | 16070 (18.34) | **<.001** |
| | N (Row %) | Black | 8935 (78.15) | 2498 (21.85) | |
| | N (Row %) | other | 3249 (79.3) | 848 (20.7) | |
| Sex | N (Row %) | Male | 38587 (79.9) | 9707 (20.1) | **<.001** |
| | N (Row %) | Female | 45137 (82.3) | 9709 (17.7) | |
| Spanish Hispanic Origin | N (Row %) | Non-Spanish; non-Hispanic | 73990 (81.73) | 16540 (18.27) | **<.001** |
| | N (Row %) | Spanish or Hispanic | 3547 (74.05) | 1243 (25.95) | |
| | N (Row %) | Unknown | 6187 (79.12) | 1633 (20.88) | |
| Year of Diagnosis | N (Row %) | 2004 | 9499 (79.64) | 2428 (20.36) | **<.001** |
| | N (Row %) | 2005 | 9412 (79.86) | 2374 (20.14) | |
| | N (Row %) | 2006 | 9353 (78.88) | 2505 (21.12) | |
| | N (Row %) | 2007 | 9189 (79.81) | 2324 (20.19) | |
| | N (Row %) | 2008 | 9227 (81.4) | 2108 (18.6) | |
| | N (Row %) | 2009 | 9092 (81.72) | 2034 (18.28) | |
| | N (Row %) | 2010 | 9054 (83.34) | 1810 (16.66) | |
| | N (Row %) | 2011 | 9368 (82.95) | 1925 (17.05) | |
| | N (Row %) | 2012 | 9530 (83.32) | 1908 (16.68) | |
| Primary Payer | N (Row %) | Not Insured | 2255 (66.62) | 1130 (33.38) | **<.001** |
| | N (Row %) | Private Insurance | 21058 (68.03) | 9894 (31.97) | |
| | N (Row %) | Medicaid | 2544 (69.53) | 1115 (30.47) | |
| | N (Row %) | Medicare | 56168 (89.18) | 6812 (10.82) | |
| | N (Row %) | Other Government | 494 (75.19) | 163 (24.81) | |
| | N (Row %) | Insurance Status Unknown | 1205 (79.96) | 302 (20.04) | |

| Covariate | Statistics | Level | Chemotherapy at any CoC Facility | | Parametric P-value* |
|---|---|---|---|---|---|
| | | | No N=83724 | Yes N=19416 | |
| Facility Type | N (Row %) | Community Cancer Program/Other | 12815 (80.93) | 3019 (19.07) | **<.001** |
| | N (Row %) | Comprehensive Community Cancer Program | 43855 (82.49) | 9310 (17.51) | |
| | N (Row %) | Academic/Research Program | 19818 (81.02) | 4642 (18.98) | |
| | N (Row %) | Integrated Network Cancer Program | 6308 (82.2) | 1366 (17.8) | |
| Facility Location | N (Row %) | Northeast | 16330 (81.1) | 3805 (18.9) | **<.001** |
| | N (Row %) | South | 30913 (81.52) | 7008 (18.48) | |
| | N (Row %) | Midwest | 23150 (81.67) | 5196 (18.33) | |
| | N (Row %) | West | 12403 (84.2) | 2328 (15.8) | |
| Median Income Quartiles 2000 | N (Row %) | < $30,000 | 11110 (80.02) | 2774 (19.98) | **<.001** |
| | N (Row %) | $30,000 - $35,999 | 14877 (81.06) | 3477 (18.94) | |
| | N (Row %) | $36,000 - $45,999 | 23350 (81.77) | 5206 (18.23) | |
| | N (Row %) | $46,000 + | 31549 (81.15) | 7326 (18.85) | |
| Percent No High School Degree Quartiles 2000 | N (Row %) | >=29% | 13494 (79.9) | 3395 (20.1) | **<.001** |
| | N (Row %) | 20-28.9% | 19012 (80.45) | 4621 (19.55) | |
| | N (Row %) | 14-19.9% | 19662 (81.68) | 4409 (18.32) | |
| | N (Row %) | < 14% | 28713 (81.88) | 6356 (18.12) | |
| Urban/Rural 2003 | N (Row %) | 1-Metro | 67540 (81.22) | 15614 (18.78) | 0.258 |
| | N (Row %) | 2-Urban | 11705 (80.65) | 2809 (19.35) | |
| | N (Row %) | 3-Rural | 1765 (81) | 414 (19) | |
| Sequence Number | N (Row %) | 0 | 74832 (81.08) | 17465 (18.92) | **0.019** |
| | N (Row %) | 1 | 8892 (82.01) | 1951 (17.99) | |

| Covariate | Statistics | Level | Chemotherapy at any CoC Facility | | Parametric P-value* |
|---|---|---|---|---|---|
| | | | No N=83724 | Yes N=19416 | |
| Charlson-Deyo Score | N (Row %) | 0 | 54155 (78.34) | 14975 (21.66) | **<.001** |
| | N (Row %) | 1 | 20714 (85.2) | 3599 (14.8) | |
| | N (Row %) | 2+ | 8855 (91.32) | 842 (8.68) | |
| Grade | N (Row %) | Well differentiated, differentiated, NOS | 8099 (82.28) | 1744 (17.72) | **<.001** |
| | N (Row %) | Moderately differentiated, moderately well differentiated, intermediate differentiation | 60497 (82.09) | 13203 (17.91) | |
| | N (Row %) | Poorly differentiated | 12178 (77.45) | 3546 (22.55) | |
| | N (Row %) | Undifferentiated, anaplastic | 1332 (78.31) | 369 (21.69) | |
| | N (Row %) | Cell type not determined, not stated or not applicable, unknown primaries, high grade dysplasia | 1618 (74.49) | 554 (25.51) | |
| Surgical Margins Status at any CoC Facility | N (Row %) | no | 80764 (81.79) | 17978 (18.21) | **<.001** |
| | N (Row %) | yes | 2413 (66.84) | 1197 (33.16) | |
| | N (Row %) | unknown | 547 (69.42) | 241 (30.58) | |
| Primary Site | N (Row %) | C180 Cecum | 19004 (83.16) | 3849 (16.84) | **<.001** |
| | N (Row %) | C181 Appendix | 1050 (62.35) | 634 (37.65) | |
| | N (Row %) | C182 Ascending colon | 20747 (85.42) | 3540 (14.58) | |
| | N (Row %) | C183 Hepatic flexure of colon | 5029 (84.01) | 957 (15.99) | |
| | N (Row %) | C184 Transverse colon | 9190 (81.63) | 2068 (18.37) | |
| | N (Row %) | C185 Splenic flexure of colon | 3164 (78.12) | 886 (21.88) | |
| | N (Row %) | C186 Descending colon | 4859 (76.71) | 1475 (23.29) | |
| | N (Row %) | C187 Sigmoid colon | 18476 (77.37) | 5403 (22.63) | |
| | N (Row %) | C188 Overlapping lesion of colon | 1076 (80.18) | 266 (19.82) | |
| | N (Row %) | C189 Colon, NOS | 1129 (76.96) | 338 (23.04) | |

| Covariate | Statistics | Level | Chemotherapy at any CoC Facility: No N=83724 | Chemotherapy at any CoC Facility: Yes N=19416 | Parametric P-value* |
|---|---|---|---|---|---|
| Surgical Approach at this Facility | N (Row %) | No surgical procedure of primary site | 740 (50.03) | 739 (49.97) | **<.001** |
| | N (Row %) | Robotic assisted | 479 (87.41) | 69 (12.59) | |
| | N (Row %) | Robotic converted to open | 50 (79.37) | 13 (20.63) | |
| | N (Row %) | Laparoscopic | 9321 (86.73) | 1426 (13.27) | |
| | N (Row %) | Laparoscopic converted to open | 1598 (83.23) | 322 (16.77) | |
| | N (Row %) | Open or approach unspecified | 15764 (83.68) | 3074 (16.32) | |
| AJCC Pathologic Stage Group | N (Row %) | 2 | 1705 (81.85) | 378 (18.15) | **<.001** |
| | N (Row %) | 2A | 74455 (83.93) | 14256 (16.07) | |
| | N (Row %) | 2B | 6571 (62.43) | 3955 (37.57) | |
| | N (Row %) | 2C | 993 (54.56) | 827 (45.44) | |
| Lymph Vascular Invasion | N (Row %) | Not present | 22766 (85.13) | 3976 (14.87) | **<.001** |
| | N (Row %) | present | 3245 (74.34) | 1120 (25.66) | |
| | N (Row %) | Not applicable or Unknown | 1941 (78.01) | 547 (21.99) | |
| Age at Diagnosis | N | | 83724 | 19416 | **<.001** |
| | Mean | | 72.38 | 59.75 | |
| | Median | | 75 | 60 | |
| | Min | | 18 | 18 | |
| | Max | | 90 | 90 | |
| | Std Dev | | 12.48 | 12.25 | |
| Tumor Size | N | | 81261 | 18452 | **<.001** |
| | Mean | | 5.29 | 5.79 | |
| | Median | | 4.7 | 5 | |
| | Min | | 0.1 | 0.1 | |
| | Max | | 98.9 | 98.9 | |
| | Std Dev | | 4.26 | 4.34 | |

| Covariate | Statistics | Level | Chemotherapy at any CoC Facility | | Parametric P-value* |
|---|---|---|---|---|---|
| | | | No N=83724 | Yes N=19416 | |
| Great Circle Distance | N | | 82501 | 19186 | **0.001** |
| | Mean | | 0.41 | 0.37 | |
| | Median | | 0.14 | 0.15 | |
| | Min | | 0 | 0 | |
| | Max | | 88.41 | 53.41 | |
| | Std Dev | | 1.67 | 1.36 | |

* The parametric p-value is calculated by ANOVA for numerical covariates and chi-square test for categorical covariates.

Table 3-2. Univariate Survival Analysis for Chemotherapy

| | | Overall Survival (Months) | | |
|---|---|---|---|---|
| **Covariate** | **Level** | **Hazard Ratio** | **HR P-value** | **Type3 P-value** |
| Chemotherapy | Yes | 0.40 (0.38-0.41) | **<.001** | **<.001** |
| | No | - | - | |

* Number of observations in the original data set = 103140. Number of observations used = 103079.

Table 3-3. Univariate Association with Chemo For Aim 1

| | | | Overall Survival (Months) | | |
|---|---|---|---|---|---|
| **Covariate** | **Level** | **N** | **Hazard Ratio (95% CI)** | **HR P-value** | **Log-rank P-value** |
| Race | other | 4097 | 0.66 (0.62-0.71) | **<.001** | **<.001** |
| | Black | 11433 | 0.90 (0.87-0.93) | **<.001** | |
| | White | 87610 | - | - | |
| Sex | Male | 48294 | 1.04 (1.02-1.06) | **<.001** | **<.001** |
| | Female | 54846 | - | - | |
| Spanish Hispanic Origin | Unknown | 7820 | 1.00 (0.96-1.04) | 0.936 | **<.001** |
| | Spanish or Hispanic | 4790 | 0.71 (0.66-0.75) | **<.001** | |
| | Non-Spanish; non-Hispanic | 90530 | - | - | |
| Year of Diagnosis | 2004 | 11927 | 1.10 (1.03-1.16) | **0.002** | **<.001** |
| | 2005 | 11786 | 1.10 (1.04-1.17) | **0.001** | |
| | 2006 | 11858 | 1.12 (1.06-1.19) | **<.001** | |
| | 2007 | 11513 | 1.10 (1.03-1.16) | **0.003** | |
| | 2008 | 11335 | 1.08 (1.02-1.15) | **0.012** | |
| | 2009 | 11126 | 1.06 (1.00-1.13) | 0.069 | |
| | 2010 | 10864 | 1.09 (1.02-1.16) | **0.008** | |
| | 2011 | 11293 | 0.99 (0.93-1.06) | 0.843 | |
| | 2012 | 11438 | - | - | |
| Primary Payer | Insurance Status Unknown | 1507 | 1.60 (1.42-1.81) | **<.001** | **<.001** |
| | Other Government | 657 | 1.30 (1.08-1.55) | **0.005** | |
| | Medicare | 62980 | 2.41 (2.22-2.61) | **<.001** | |
| | Medicaid | 3659 | 1.57 (1.42-1.74) | **<.001** | |
| | Private Insurance | 30952 | 0.84 (0.78-0.92) | **<.001** | |
| | Not Insured | 3385 | - | - | |
| Facility Type | Community Cancer Program/Other | 15834 | 1.13 (1.08-1.18) | **<.001** | **<.001** |
| | Comprehensive Community Cancer Program | 53165 | 1.04 (0.99-1.08) | 0.093 | |
| | Academic/Research Program | 24460 | 0.90 (0.86-0.94) | **<.001** | |
| | Integrated Network Cancer Program | 7674 | - | - | |

| | | | Overall Survival (Months) | | |
|---|---|---|---|---|---|
| | | | ---------- | | |
| **Covariate** | **Level** | **N** | **Hazard Ratio (95% CI)** | **HR P-value** | **Log-rank P-value** |
| Facility Location | Northeast | 20135 | 1.15 (1.11-1.19) | **<.001** | **<.001** |
| | South | 37921 | 1.05 (1.01-1.09) | **0.005** | |
| | Midwest | 28346 | 1.12 (1.08-1.16) | **<.001** | |
| | West | 14731 | - | **-** | |
| Median Income Quartiles 2000 | < $30,000 | 13884 | 1.18 (1.14-1.22) | **<.001** | **<.001** |
| | $30,000 - $35,999 | 18354 | 1.14 (1.11-1.18) | **<.001** | |
| | $36,000 - $45,999 | 28556 | 1.08 (1.05-1.11) | **<.001** | |
| | $46,000 + | 38875 | - | **-** | |
| Percent No High School Degree Quartiles 2000 | >=29% | 16889 | 1.07 (1.04-1.11) | **<.001** | **<.001** |
| | 20-28.9% | 23633 | 1.09 (1.06-1.13) | **<.001** | |
| | 14-19.9% | 24071 | 1.11 (1.08-1.14) | **<.001** | |
| | < 14% | 35069 | - | **-** | |
| Urban/Rural 2003 | 1-Metro | 83154 | 0.96 (0.89-1.03) | 0.261 | **0.021** |
| | 2-Urban | 14514 | 1.00 (0.93-1.08) | 0.994 | |
| | 3-Rural | 2179 | - | **-** | |
| Sequence Number | 0 | 92297 | 0.81 (0.78-0.83) | **<.001** | **<.001** |
| | 1 | 10843 | - | **-** | |
| Charlson-Deyo Score | 0 | 69130 | 0.39 (0.38-0.41) | **<.001** | **<.001** |
| | 1 | 24313 | 0.61 (0.59-0.63) | **<.001** | |
| | 2+ | 9697 | - | **-** | |
| Grade | Well differentiated, differentiated, NOS | 9843 | 0.95 (0.88-1.04) | 0.270 | **<.001** |
| | Moderately differentiated, moderately well differentiated, intermediate differentiation | 73700 | 0.99 (0.92-1.07) | 0.769 | |
| | Poorly differentiated | 15724 | 1.18 (1.09-1.28) | **<.001** | |
| | Undifferentiated, anaplastic | 1701 | 1.26 (1.13-1.41) | **<.001** | |
| | Cell type not determined, not stated or not applicable, unknown primaries, high grade dysplasia | 2172 | - | **-** | |
| Surgical Margins Status at any CoC Facility | unknown | 788 | 1.40 (1.25-1.56) | **<.001** | **<.001** |
| | yes | 3610 | 1.91 (1.82-2.00) | **<.001** | |
| | no | 98742 | - | **-** | |

| | | | Overall Survival (Months) | | |
|---|---|---|---|---|---|
| **Covariate** | **Level** | **N** | **Hazard Ratio (95% CI)** | **HR P-value** | **Log-rank P-value** |
| Primary Site | C180 Cecum | 22853 | 0.88 (0.80-0.96) | **0.003** | **<.001** |
| | C181 Appendix | 1684 | 0.64 (0.56-0.73) | **<.001** | |
| | C182 Ascending colon | 24287 | 0.87 (0.80-0.95) | **0.002** | |
| | C183 Hepatic flexure of colon | 5986 | 0.88 (0.80-0.97) | **0.009** | |
| | C184 Transverse colon | 11258 | 0.90 (0.82-0.98) | **0.022** | |
| | C185 Splenic flexure of colon | 4050 | 0.87 (0.79-0.96) | **0.006** | |
| | C186 Descending colon | 6334 | 0.83 (0.76-0.91) | **<.001** | |
| | C187 Sigmoid colon | 23879 | 0.80 (0.73-0.88) | **<.001** | |
| | C188 Overlapping lesion of colon | 1342 | 0.98 (0.87-1.12) | 0.806 | |
| | C189 Colon, NOS | 1467 | - | - | |
| Surgical Approach at this Facility | No surgical procedure of primary site | 1479 | 0.50 (0.42-0.58) | **<.001** | **<.001** |
| | Robotic assisted | 548 | 0.41 (0.31-0.55) | **<.001** | |
| | Robotic converted to open | 63 | 0.55 (0.26-1.16) | 0.114 | |
| | Laparoscopic | 10747 | 0.59 (0.56-0.63) | **<.001** | |
| | Laparoscopic converted to open | 1920 | 0.84 (0.76-0.94) | **0.002** | |
| | Open or approach unspecified | 18838 | - | - | |
| AJCC Pathologic Stage Group | 2 | 2083 | 0.49 (0.44-0.56) | **<.001** | **<.001** |
| | 2A | 88711 | 0.60 (0.55-0.65) | **<.001** | |
| | 2B | 10526 | 0.95 (0.87-1.04) | 0.277 | |
| | 2C | 1820 | - | - | |
| Lymph Vascular Invasion | Not present | 26742 | 0.90 (0.82-0.99) | **0.025** | **<.001** |
| | present | 4365 | 1.18 (1.05-1.31) | **0.004** | |
| | Not applicable or Unknown | 2488 | - | - | |
| Age at Diagnosis | | 103079 | 1.06 (1.06-1.06) | **<.001** | - |
| Tumor Size | | 99656 | 1.00 (1.00-1.00) | 0.099 | - |
| Great Circle Distance | | 101629 | 1.00 (0.99-1.01) | 0.510 | - |

Table 3-4. Univariate Association with Survival For Aim 2- Overall survival

| | | | Overall Survival (Months) | | |
|---|---|---|---|---|---|
| **Covariate** | **Level** | **N** | **Hazard Ratio (95% CI)** | **HR P-value** | **Log-rank P-value** |
| Race | other | 848 | 0.62 (0.50-0.77) | **<.001** | **<.001** |
| | Black | 2498 | 1.23 (1.12-1.35) | **<.001** | |
| | White | 16070 | - | - | |
| Sex | Male | 9707 | 1.17 (1.09-1.25) | **<.001** | **<.001** |
| | Female | 9709 | - | - | |
| Spanish Hispanic Origin | Unknown | 1633 | 1.04 (0.93-1.16) | 0.528 | **0.019** |
| | Spanish or Hispanic | 1243 | 0.81 (0.69-0.94) | **0.007** | |
| | Non-Spanish; non-Hispanic | 16540 | - | - | |
| Year of Diagnosis | 2004 | 2428 | 0.93 (0.76-1.15) | 0.506 | 0.171 |
| | 2005 | 2374 | 0.83 (0.68-1.03) | 0.091 | |
| | 2006 | 2505 | 0.93 (0.76-1.15) | 0.523 | |
| | 2007 | 2324 | 0.91 (0.74-1.13) | 0.395 | |
| | 2008 | 2108 | 1.02 (0.82-1.26) | 0.854 | |
| | 2009 | 2034 | 0.95 (0.76-1.19) | 0.667 | |
| | 2010 | 1810 | 0.96 (0.76-1.21) | 0.747 | |
| | 2011 | 1925 | 1.00 (0.79-1.27) | 0.976 | |
| | 2012 | 1908 | - | - | |
| Primary Payer | Insurance Status Unknown | 302 | 1.08 (0.81-1.44) | 0.618 | **<.001** |
| | Other Government | 163 | 0.81 (0.52-1.26) | 0.357 | |
| | Medicare | 6812 | 1.51 (1.30-1.75) | **<.001** | |
| | Medicaid | 1115 | 1.23 (1.01-1.49) | **0.040** | |
| | Private Insurance | 9894 | 0.60 (0.52-0.70) | **<.001** | |
| | Not Insured | 1130 | - | - | |
| Facility Type | Community Cancer Program/Other | 3019 | 1.15 (1.00-1.33) | **0.050** | **<.001** |
| | Comprehensive Community Cancer Program | 9310 | 1.00 (0.88-1.14) | 0.985 | |
| | Academic/Research Program | 4642 | 0.88 (0.76-1.01) | 0.078 | |
| | Integrated Network Cancer Program | 1366 | - | - | |

| | | | Overall Survival (Months) | | |
|---|---|---|---|---|---|
| **Covariate** | **Level** | **N** | **Hazard Ratio (95% CI)** | **HR P-value** | **Log-rank P-value** |
| Facility Location | Northeast | 3805 | 1.27 (1.12-1.43) | **<.001** | **<.001** |
| | South | 7008 | 1.24 (1.11-1.39) | **<.001** | |
| | Midwest | 5196 | 1.25 (1.11-1.40) | **<.001** | |
| | West | 2328 | - | - | |
| Median Income Quartiles 2000 | < $30,000 | 2774 | 1.58 (1.44-1.75) | **<.001** | **<.001** |
| | $30,000 - $35,999 | 3477 | 1.37 (1.25-1.51) | **<.001** | |
| | $36,000 - $45,999 | 5206 | 1.25 (1.15-1.37) | **<.001** | |
| | $46,000 + | 7326 | - | - | |
| Percent No High School Degree Quartiles 2000 | >=29% | 3395 | 1.47 (1.33-1.62) | **<.001** | **<.001** |
| | 20-28.9% | 4621 | 1.33 (1.21-1.45) | **<.001** | |
| | 14-19.9% | 4409 | 1.32 (1.20-1.44) | **<.001** | |
| | < 14% | 6356 | - | - | |
| Urban/Rural 2003 | 1-Metro | 15614 | 0.88 (0.71-1.10) | 0.261 | 0.094 |
| | 2-Urban | 2809 | 0.97 (0.77-1.22) | 0.769 | |
| | 3-Rural | 414 | - | - | |
| Sequence Number | 0 | 17465 | 0.64 (0.59-0.70) | **<.001** | **<.001** |
| | 1 | 1951 | - | - | |
| Charlson-Deyo Score | 0 | 14975 | 0.40 (0.36-0.46) | **<.001** | **<.001** |
| | 1 | 3599 | 0.63 (0.55-0.72) | **<.001** | |
| | 2+ | 842 | - | - | |
| Grade | Well differentiated, differentiated, NOS | 1744 | 0.65 (0.53-0.81) | **<.001** | **<.001** |
| | Moderately differentiated, moderately well differentiated, intermediate differentiation | 13203 | 0.73 (0.61-0.88) | **0.001** | |
| | Poorly differentiated | 3546 | 0.89 (0.73-1.08) | 0.230 | |
| | Undifferentiated, anaplastic | 369 | 1.16 (0.88-1.54) | 0.290 | |
| | Cell type not determined, not stated or not applicable, unknown primaries, high grade dysplasia | 554 | - | - | |
| Surgical Margins Status at any CoC Facility | unknown | 241 | 1.85 (1.45-2.37) | **<.001** | **<.001** |
| | yes | 1197 | 2.41 (2.17-2.67) | **<.001** | |
| | no | 17978 | - | - | |
| Age at Diagnosis | | 19415 | 1.04 (1.04-1.05) | **<.001** | - |
| Tumor Size | | 18451 | 1.00 (0.99-1.01) | 0.510 | - |

| | | | Overall Survival (Months) | | |
|---|---|---|---|---|---|
| | | | --- | | |
| **Covariate** | **Level** | **N** | **Hazard Ratio (95% CI)** | **HR P-value** | **Log-rank P-value** |
| Great Circle Distance | | 19186 | 1.02 (1.00-1.04) | 0.081 | - |

Table 4-1. Multivariable Logistic Regression for Aim 1

| | | Chemotherapy at any CoC Facility=Yes | | |
|---|---|---|---|---|
| **Covariate** | **Level** | **Odds Ratio (95% CI)** | **OR P-value** | **Type3 P-value** |
| Race | other | 0.88 (0.80-0.97) | **0.008** | **<.001** |
| | Black | 0.87 (0.82-0.92) | **<.001** | |
| | White | - | - | |
| Spanish Hispanic Origin | Unknown | 1.12 (1.05-1.20) | **<.001** | **<.001** |
| | Spanish or Hispanic | 1.13 (1.04-1.22) | **0.005** | |
| | Non-Spanish; non-Hispanic | - | - | |
| Year of Diagnosis | 2004 | 1.50 (1.39-1.62) | **<.001** | **<.001** |
| | 2005 | 1.42 (1.31-1.53) | **<.001** | |
| | 2006 | 1.53 (1.41-1.65) | **<.001** | |
| | 2007 | 1.43 (1.32-1.55) | **<.001** | |
| | 2008 | 1.24 (1.14-1.34) | **<.001** | |
| | 2009 | 1.17 (1.08-1.26) | **<.001** | |
| | 2010 | 1.04 (0.95-1.12) | 0.402 | |
| | 2011 | 1.04 (0.96-1.12) | 0.396 | |
| | 2012 | - | - | |
| Primary Payer | Insurance Status Unknown | 0.86 (0.72-1.02) | 0.080 | **<.001** |
| | Other Government | 1.02 (0.82-1.28) | 0.849 | |
| | Medicare | 1.11 (1.01-1.22) | **0.037** | |
| | Medicaid | 1.06 (0.94-1.19) | 0.333 | |
| | Private Insurance | 1.19 (1.09-1.30) | **<.001** | |
| | Not Insured | - | - | |
| Facility Type | Community Cancer Program/Other | 1.14 (1.05-1.23) | **0.002** | **<.001** |
| | Comprehensive Community Cancer Program | 1.02 (0.95-1.10) | 0.582 | |
| | Academic/Research Program | 0.90 (0.84-0.98) | **0.009** | |
| | Integrated Network Cancer Program | - | - | |
| Facility Location | Northeast | 1.45 (1.36-1.55) | **<.001** | **<.001** |
| | South | 1.19 (1.12-1.26) | **<.001** | |
| | Midwest | 1.35 (1.26-1.43) | **<.001** | |
| | West | - | - | |

| Covariate | Level | Chemotherapy at any CoC Facility=Yes<br>Odds Ratio (95% CI) | OR P-value | Type3 P-value |
|---|---|---|---|---|
| Median Income Quartiles 2000 | < $30,000 | 1.10 (1.04-1.17) | **0.001** | **0.002** |
| | $30,000 - $35,999 | 1.03 (0.98-1.09) | 0.274 | |
| | $36,000 - $45,999 | 0.98 (0.94-1.03) | 0.447 | |
| | $46,000 + | - | - | |
| Sequence Number | 0 | 0.90 (0.84-0.95) | **<.001** | **<.001** |
| | 1 | - | - | |
| Charlson-Deyo Score | 0 | 1.85 (1.71-2.00) | **<.001** | **<.001** |
| | 1 | 1.57 (1.44-1.72) | **<.001** | |
| | 2+ | - | - | |
| Grade | Well differentiated, differentiated, NOS | 0.75 (0.66-0.87) | **<.001** | **<.001** |
| | Moderately differentiated, moderately well differentiated, intermediate differentiation | 0.80 (0.71-0.91) | **<.001** | |
| | Poorly differentiated | 1.29 (1.13-1.48) | **<.001** | |
| | Undifferentiated, anaplastic | 1.34 (1.11-1.61) | **0.002** | |
| | Cell type not determined, not stated or not applicable, unknown primaries, high grade dysplasia | - | - | |
| Surgical Margins Status at any CoC Facility | unknown | 1.42 (1.14-1.76) | **0.002** | **<.001** |
| | yes | 2.63 (2.42-2.86) | **<.001** | |
| | no | - | - | |
| Age at Diagnosis | | 0.92 (0.92-0.93) | **<.001** | **<.001** |
| Tumor Size | | 1.02 (1.01-1.02) | **<.001** | **<.001** |
| Great Circle Distance | | 0.96 (0.95-0.98) | **<.001** | **<.001** |

* Number of observations in the original data set = 103140. Number of observations used = 94553.
** Backward selection with an alpha level of removal of .10 was used. The following variables were removed from the model: Percent No High School Degree Quartiles 2000, Sex, and Urban/Rural 2003.

Table 4-2. Multivariable Survival Analysis for Aim 1

| | | Overall Survival (Months) | | |
|---|---|---|---|---|
| **Covariate** | **Level** | **Hazard Ratio** | **HR P-value** | **Type3 P-value** |
| Chemotherapy at any CoC Facility | Yes | 0.83 (0.80-0.87) | **<.001** | **<.001** |
| | No | - | - | |
| Age at Diagnosis | | 1.06 (1.06-1.06) | **<.001** | **<.001** |
| Race | White | 1.20 (1.12-1.30) | **<.001** | **<.001** |
| | Black | 1.36 (1.25-1.48) | **<.001** | |
| | other | - | - | |
| Sex | Male | 1.25 (1.22-1.28) | **<.001** | **<.001** |
| | Female | - | - | |
| Spanish Hispanic Origin | Non-Spanish; non-Hispanic | 1.00 (0.96-1.04) | 0.938 | **0.001** |
| | Spanish or Hispanic | 0.87 (0.80-0.94) | **<.001** | |
| | Unknown | - | - | |
| Primary Payer | Not Insured | 1.32 (1.15-1.53) | **<.001** | **<.001** |
| | Private Insurance | 0.86 (0.77-0.96) | **0.007** | |
| | Medicaid | 1.60 (1.41-1.83) | **<.001** | |
| | Medicare | 0.96 (0.86-1.07) | 0.432 | |
| | Other Government | 1.07 (0.87-1.33) | 0.511 | |
| | Insurance Status Unknown | - | - | |
| Facility Type | Community Cancer Program/Other | 1.09 (1.03-1.15) | **0.002** | **<.001** |
| | Comprehensive Community Cancer Program | 1.03 (0.98-1.08) | 0.236 | |

| | | Overall Survival (Months) | | |
|---|---|---|---|---|
| **Covariate** | **Level** | **Hazard Ratio** | **HR P-value** | **Type3 P-value** |
| | Academic/Research Program | 0.98 (0.93-1.03) | 0.497 | |
| | Integrated Network Cancer Program | - | - | |
| Median Income Quartiles 2000 | < $30,000 | 1.16 (1.10-1.22) | **<.001** | **<.001** |
| | $30,000 - $35,999 | 1.09 (1.04-1.13) | **<.001** | |
| | $36,000 - $45,999 | 1.03 (0.99-1.07) | 0.103 | |
| | $46,000 + | - | - | |
| Percent No High School Degree Quartiles 2000 | >=29% | 1.07 (1.02-1.13) | **0.006** | **0.003** |
| | 20-28.9% | 1.07 (1.03-1.11) | **<.001** | |
| | 14-19.9% | 1.05 (1.02-1.09) | **0.008** | |
| | < 14% | - | - | |
| Urban/Rural 2003 | 1-Metro | 1.10 (1.01-1.19) | **0.031** | 0.073 |
| | 2-Urban | 1.07 (0.99-1.17) | 0.103 | |
| | 3-Rural | - | - | |
| Sequence Number | 0 | 0.77 (0.75-0.80) | **<.001** | **<.001** |
| | 1 | - | - | |
| Charlson-Deyo Score | 0 | 0.50 (0.48-0.51) | **<.001** | **<.001** |
| | 1 | 0.66 (0.63-0.68) | **<.001** | |
| | 2+ | - | - | |

| | | Overall Survival (Months) | | |
|---|---|---|---|---|
| | | ---------- | | |
| **Covariate** | **Level** | **Hazard Ratio** | **HR P-value** | **Type3 P-value** |
| Grade | Well differentiated, differentiated, NOS | 0.90 (0.81-0.99) | **0.032** | **<.001** |
| | Moderately differentiated, moderately well differentiated, intermediate differentiation | 0.92 (0.84-1.01) | 0.079 | |
| | Poorly differentiated | 1.02 (0.93-1.12) | 0.798 | |
| | Undifferentiated, anaplastic | 1.05 (0.92-1.19) | 0.366 | |
| | Cell type not determined, not stated or not applicable, unknown primaries, high grade dysplasia | - | - | |
| Surgical Margins Status at any CoC Facility | no | 0.69 (0.61-0.79) | **<.001** | **<.001** |
| | yes | 1.35 (1.17-1.55) | **<.001** | |
| | unknown | - | - | |

* Number of observations in the original data set = 97033. Number of observations used = 88557.
** Backward selection with an alpha level of removal of .10 was used. The following variables were removed from the model: Great Circle Distance, and Facility Location, Year of Diagnosis, and Tumor Size.

Table 4-3. Multivariable Survival Analysis for Aim 2- Overall survival

| | | Overall Survival (Months) | | |
|---|---|---|---|---|
| **Covariate** | **Level** | **Hazard Ratio** | **HR P-value** | **Type3 P-value** |
| Race | White | 1.31 (1.06-1.64) | **0.015** | **<.001** |
| | Black | 1.61 (1.27-2.04) | **<.001** | |
| | other | - | - | |
| Sex | Male | 1.25 (1.17-1.34) | **<.001** | **<.001** |
| | Female | - | - | |
| Primary Payer | Not Insured | 1.02 (0.75-1.39) | 0.883 | **<.001** |
| | Private Insurance | 0.61 (0.47-0.81) | **<.001** | |
| | Medicaid | 1.20 (0.89-1.63) | 0.229 | |
| | Medicare | 0.78 (0.59-1.03) | 0.083 | |
| | Other Government | 0.61 (0.35-1.07) | 0.085 | |
| | Insurance Status Unknown | - | - | |
| Facility Type | Community Cancer Program/Other | 1.07 (0.92-1.25) | 0.352 | **0.018** |
| | Comprehensive Community Cancer Program | 0.97 (0.85-1.11) | 0.678 | |
| | Academic/Research Program | 0.90 (0.78-1.04) | 0.158 | |
| | Integrated Network Cancer Program | - | - | |
| Median Income Quartiles 2000 | < $30,000 | 1.36 (1.18-1.56) | **<.001** | **<.001** |
| | $30,000 - $35,999 | 1.20 (1.06-1.36) | **0.003** | |
| | $36,000 - $45,999 | 1.10 (1.00-1.22) | 0.056 | |
| | $46,000 + | - | - | |
| Percent No High School Degree Quartiles 2000 | >=29% | 1.12 (0.98-1.29) | 0.094 | 0.071 |
| | 20-28.9% | 1.12 (1.00-1.26) | 0.051 | |
| | 14-19.9% | 1.15 (1.03-1.28) | **0.009** | |
| | < 14% | - | - | |
| Urban/Rural 2003 | 1-Metro | 1.22 (0.97-1.54) | 0.084 | 0.074 |
| | 2-Urban | 1.12 (0.88-1.41) | 0.358 | |
| | 3-Rural | - | - | |
| Sequence Number | 0 | 0.76 (0.69-0.83) | **<.001** | **<.001** |
| | 1 | - | - | |

| | | Overall Survival (Months) | | |
|---|---|---|---|---|
| **Covariate** | **Level** | **Hazard Ratio** | **HR P-value** | **Type3 P-value** |
| Charlson-Deyo Score | 0 | 0.54 (0.48-0.62) | **<.001** | **<.001** |
| | 1 | 0.69 (0.60-0.79) | **<.001** | |
| | 2+ | - | - | |
| Grade | Well differentiated, differentiated, NOS | 0.67 (0.54-0.85) | **<.001** | **<.001** |
| | Moderately differentiated, moderately well differentiated, intermediate differentiation | 0.75 (0.61-0.91) | **0.003** | |
| | Poorly differentiated | 0.83 (0.68-1.02) | 0.084 | |
| | Undifferentiated, anaplastic | 1.04 (0.77-1.40) | 0.806 | |
| | Cell type not determined, not stated or not applicable, unknown primaries, high grade dysplasia | - | - | |
| Surgical Margins Status at any CoC Facility | no | 0.53 (0.40-0.69) | **<.001** | **<.001** |
| | yes | 1.16 (0.87-1.54) | 0.322 | |
| | unknown | - | - | |
| Age at Diagnosis | | 1.04 (1.04-1.05) | **<.001** | **<.001** |
| Great Circle Distance | | 1.04 (1.01-1.06) | **0.002** | **0.002** |

* Number of observations in the original data set = 19416. Number of observations used = 17444.
** Backward selection with an alpha level of removal of .10 was used. The following variables were removed from the model: Facility Location, Spanish Hispanic Origin, Year of Diagnosis, and Tumor Size.

Table 5-1. PS-Balance Check after Matching

| | | | Chemotherapy | | | |
|---|---|---|---|---|---|---|
| **Covariate** | **Level** | **Statistics** | **No N=16548** | **Yes N=16548** | **Parametric P-value*** | **Standardized Difference** |
| Race | White | N (Col%) | 13664 (82.57) | 13742 (83.04) | 0.486 | 0.012 |
| | Black | N (Col%) | 2163 (13.07) | 2117 (12.79) | | 0.008 |
| | other | N (Col%) | 721 (4.36) | 689 (4.16) | | 0.010 |
| Sex | Male | N (Col%) | 8311 (50.22) | 8264 (49.94) | 0.605 | 0.006 |
| | Female | N (Col%) | 8237 (49.78) | 8284 (50.06) | | 0.006 |
| Spanish Hispanic Origin | Non-Spanish; non-Hispanic | N (Col%) | 14212 (85.88) | 14195 (85.78) | 0.949 | 0.003 |
| | Spanish or Hispanic | N (Col%) | 940 (5.68) | 953 (5.76) | | 0.003 |
| | Unknown | N (Col%) | 1396 (8.44) | 1400 (8.46) | | 0.001 |
| Year of Diagnosis | 2004 | N (Col%) | 2087 (12.61) | 2045 (12.36) | 0.867 | 0.008 |
| | 2005 | N (Col%) | 1994 (12.05) | 1963 (11.86) | | 0.006 |
| | 2006 | N (Col%) | 2056 (12.42) | 2104 (12.71) | | 0.009 |
| | 2007 | N (Col%) | 1923 (11.62) | 1979 (11.96) | | 0.010 |
| | 2008 | N (Col%) | 1764 (10.66) | 1795 (10.85) | | 0.006 |
| | 2009 | N (Col%) | 1711 (10.34) | 1744 (10.54) | | 0.007 |
| | 2010 | N (Col%) | 1582 (9.56) | 1566 (9.46) | | 0.003 |
| | 2011 | N (Col%) | 1727 (10.44) | 1666 (10.07) | | 0.012 |
| | 2012 | N (Col%) | 1704 (10.3) | 1686 (10.19) | | 0.004 |
| Primary Payer | Not Insured | N (Col%) | 919 (5.55) | 894 (5.4) | 0.705 | 0.007 |
| | Private Insurance | N (Col%) | 8028 (48.51) | 8153 (49.27) | | 0.015 |
| | Medicaid | N (Col%) | 862 (5.21) | 859 (5.19) | | 0.001 |
| | Medicare | N (Col%) | 6356 (38.41) | 6282 (37.96) | | 0.009 |
| | Other Government | N (Col%) | 148 (0.89) | 130 (0.79) | | 0.012 |
| | Insurance Status Unknown | N (Col%) | 235 (1.42) | 230 (1.39) | | 0.003 |

| Covariate | Level | Statistics | Chemotherapy | | Parametric P-value* | Standardized Difference |
|---|---|---|---|---|---|---|
| | | | No N=16548 | Yes N=16548 | | |
| Facility Type | Community Cancer Program/Other | N (Col%) | 2685 (16.23) | 2720 (16.44) | 0.643 | 0.006 |
| | Comprehensive Community Cancer Program | N (Col%) | 8386 (50.68) | 8451 (51.07) | | 0.008 |
| | Academic/Research Program | N (Col%) | 4238 (25.61) | 4138 (25.01) | | 0.014 |
| | Integrated Network Cancer Program | N (Col%) | 1239 (7.49) | 1239 (7.49) | | 0.000 |
| Facility Location | Northeast | N (Col%) | 3359 (20.3) | 3379 (20.42) | 0.581 | 0.003 |
| | South | N (Col%) | 6386 (38.59) | 6265 (37.86) | | 0.015 |
| | Midwest | N (Col%) | 4724 (28.55) | 4795 (28.98) | | 0.009 |
| | West | N (Col%) | 2079 (12.56) | 2109 (12.74) | | 0.005 |
| Median Income Quartiles 2000 | < $30,000 | N (Col%) | 2498 (15.1) | 2446 (14.78) | 0.854 | 0.009 |
| | $30,000 - $35,999 | N (Col%) | 3026 (18.29) | 3045 (18.4) | | 0.003 |
| | $36,000 - $45,999 | N (Col%) | 4631 (27.99) | 4619 (27.91) | | 0.002 |
| | $46,000 + | N (Col%) | 6393 (38.63) | 6438 (38.91) | | 0.006 |
| Sequence Number | 0 | N (Col%) | 14714 (88.92) | 14813 (89.52) | 0.079 | 0.019 |
| | 1 | N (Col%) | 1834 (11.08) | 1735 (10.48) | | 0.019 |
| Charlson-Deyo Score | 0 | N (Col%) | 12479 (75.41) | 12468 (75.34) | 0.555 | 0.002 |
| | 1 | N (Col%) | 3325 (20.09) | 3296 (19.92) | | 0.004 |
| | 2+ | N (Col%) | 744 (4.5) | 784 (4.74) | | 0.012 |

| Covariate | Level | Statistics | Chemotherapy No N=16548 | Chemotherapy Yes N=16548 | Parametric P-value* | Standardized Difference |
|---|---|---|---|---|---|---|
| Grade | Well differentiated, differentiated, NOS | N (Col%) | 1529 (9.24) | 1470 (8.88) | 0.335 | 0.012 |
| | Moderately differentiated, moderately well differentiated, intermediate differentiation | N (Col%) | 11528 (69.66) | 11450 (69.19) | | 0.010 |
| | Poorly differentiated | N (Col%) | 2839 (17.16) | 2945 (17.8) | | 0.017 |
| | Undifferentiated, anaplastic | N (Col%) | 314 (1.9) | 316 (1.91) | | 0.001 |
| | Cell type not determined, not stated or not applicable, unknown primaries, high grade dysplasia | N (Col%) | 338 (2.04) | 367 (2.22) | | 0.012 |
| Surgical Margins Status at any CoC Facility | no | N (Col%) | 15558 (94.02) | 15470 (93.49) | 0.104 | 0.022 |
| | yes | N (Col%) | 860 (5.2) | 948 (5.73) | | 0.023 |
| | unknown | N (Col%) | 130 (0.79) | 130 (0.79) | | 0.000 |
| Age at Diagnosis | | Mean (Std) | 61.51 (11.34) | 61.66 (10.6) | 0.207 | 0.014 |
| Tumor Size | | Mean (Std) | 5.65 (5.78) | 5.69 (3.87) | 0.386 | 0.010 |
| Great Circle Distance | | Mean (Std) | 0.36 (1.03) | 0.36 (1.29) | 0.765 | 0.003 |

* The parametric p value is calculated by ANOVA for numerical covariates and Chi-Square test for categorical covariates.

Table 5-2. Univariate Association with Survival for Chemo - matched

| | | Overall Survival (Months) | | |
|---|---|---|---|---|
| **Covariate** | **Level** | **Hazard Ratio** | **HR P-value** | **Type3 P-value** |
| Chemotherapy | Yes | 0.72 (0.69-0.76) | **<.001** | **<.001** |
| | No | - | - | |

* Number of observations in the original data set = 33096. Number of observations used = 33085.

Table 6. Stratified analysis - matched dataset

| Stratification | Effect Comparison | Overall Survival (Months) Hazard Ratio (95%CI) | HR P-value | Type3 P-value |
|---|---|---|---|---|
| **Race :** | **Chemotherapy at any CoC Facility :** | - | - | **0.026** |
| White | No (n=13664) vs. Yes (n=13742) | 1.42 (1.35-1.49) | **<.001** | - |
| Black | No (n=2163) vs. Yes (n=2117) | 1.17 (1.03-1.33) | **0.015** | - |
| other | No (n=721) vs. Yes (n=689) | 1.43 (1.08-1.90) | **0.011** | - |
| **Primary Payer :** | **Chemotherapy at any CoC Facility :** | - | - | **0.013** |
| Not Insured | No (n=919) vs. Yes (n=894) | 1.13 (0.90-1.41) | 0.283 | - |
| Private Insurance | No (n=8028) vs. Yes (n=8153) | 1.24 (1.13-1.35) | **<.001** | - |
| Medicaid | No (n=862) vs. Yes (n=859) | 1.47 (1.21-1.78) | **<.001** | - |
| Medicare | No (n=6356) vs. Yes (n=6282) | 1.46 (1.38-1.56) | **<.001** | - |
| Other Government | No (n=148) vs. Yes (n=130) | 1.28 (0.66-2.47) | 0.461 | - |
| Insurance Status Unknown | No (n=235) vs. Yes (n=230) | 1.73 (1.16-2.58) | **0.008** | - |
| **Median Income Quartiles 2000 :** | **Chemotherapy at any CoC Facility :** | - | - | 0.053 |
| < $30,000 | No (n=2498) vs. Yes (n=2446) | 1.23 (1.10-1.38) | **<.001** | - |
| $30,000 - $35,999 | No (n=3026) vs. Yes (n=3045) | 1.35 (1.21-1.49) | **<.001** | - |
| $36,000 - $45,999 | No (n=4631) vs. Yes (n=4619) | 1.37 (1.25-1.49) | **<.001** | - |
| $46,000 + | No (n=6393) vs. Yes (n=6438) | 1.49 (1.38-1.62) | **<.001** | - |

| Stratification | Effect Comparison | Overall Survival (Months) | | |
|---|---|---|---|---|
| | | Hazard Ratio (95%CI) | HR P-value | Type3 P-value |
| **Percent No High School Degree Quartiles 2000 :** | **Chemotherapy at any CoC Facility :** | - | - | **0.013** |
| >=29% | No (n=3079) vs. Yes (n=2961) | 1.22 (1.10-1.35) | **<.001** | - |
| 20-28.9% | No (n=3902) vs. Yes (n=4064) | 1.42 (1.29-1.56) | **<.001** | - |
| 14-19.9% | No (n=3947) vs. Yes (n=3885) | 1.32 (1.20-1.45) | **<.001** | - |
| < 14% | No (n=5618) vs. Yes (n=5636) | 1.51 (1.39-1.65) | **<.001** | - |
| **Grade :** | **Chemotherapy at any CoC Facility :** | - | - | **0.002** |
| Well differentiated, differentiated, NOS | No (n=1529) vs. Yes (n=1470) | 1.37 (1.15-1.62) | **<.001** | - |
| Moderately differentiated, moderately well differentiated, intermediate differentiation | No (n=11528) vs. Yes (n=11450) | 1.33 (1.26-1.41) | **<.001** | - |
| Poorly differentiated | No (n=2839) vs. Yes (n=2945) | 1.64 (1.48-1.82) | **<.001** | - |
| Undifferentiated, anaplastic | No (n=314) vs. Yes (n=316) | 1.21 (0.88-1.67) | 0.247 | - |
| Cell type not determined, not stated or not applicable, unknown primaries, high grade dysplasia | No (n=338) vs. Yes (n=367) | 1.01 (0.73-1.39) | 0.970 | - |
| **Surgical Margins Status at any CoC Facility :** | **Chemotherapy at any CoC Facility :** | - | - | 0.070 |

| Stratification | Effect Comparison | Overall Survival (Months) | | |
|---|---|---|---|---|
| | | Hazard Ratio (95%CI) | HR P-value | Type3 P-value |
| no | No (n=130) vs. Yes (n=130) | 1.37 (1.31-1.44) | **<.001** | - |
| yes | No (n=860) vs. Yes (n=948) | 1.64 (1.42-1.89) | **<.001** | - |
| unknown | No (n=15558) vs. Yes (n=15470) | 1.32 (0.84-2.07) | 0.227 | - |

* This table only shows interaction effect with type 3 p-value < 0.1 .

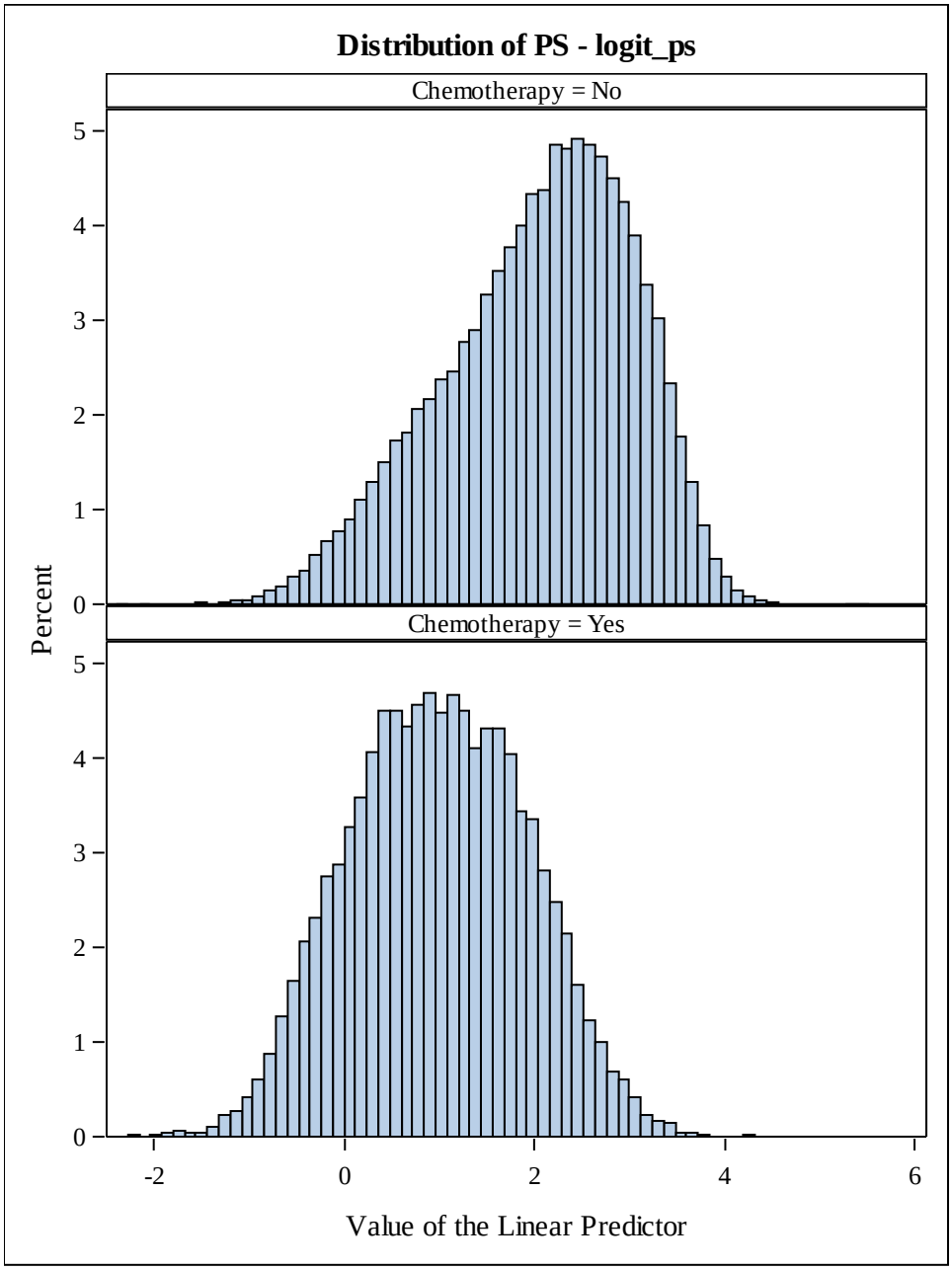


Figure 1. Distribution of propensity score matching

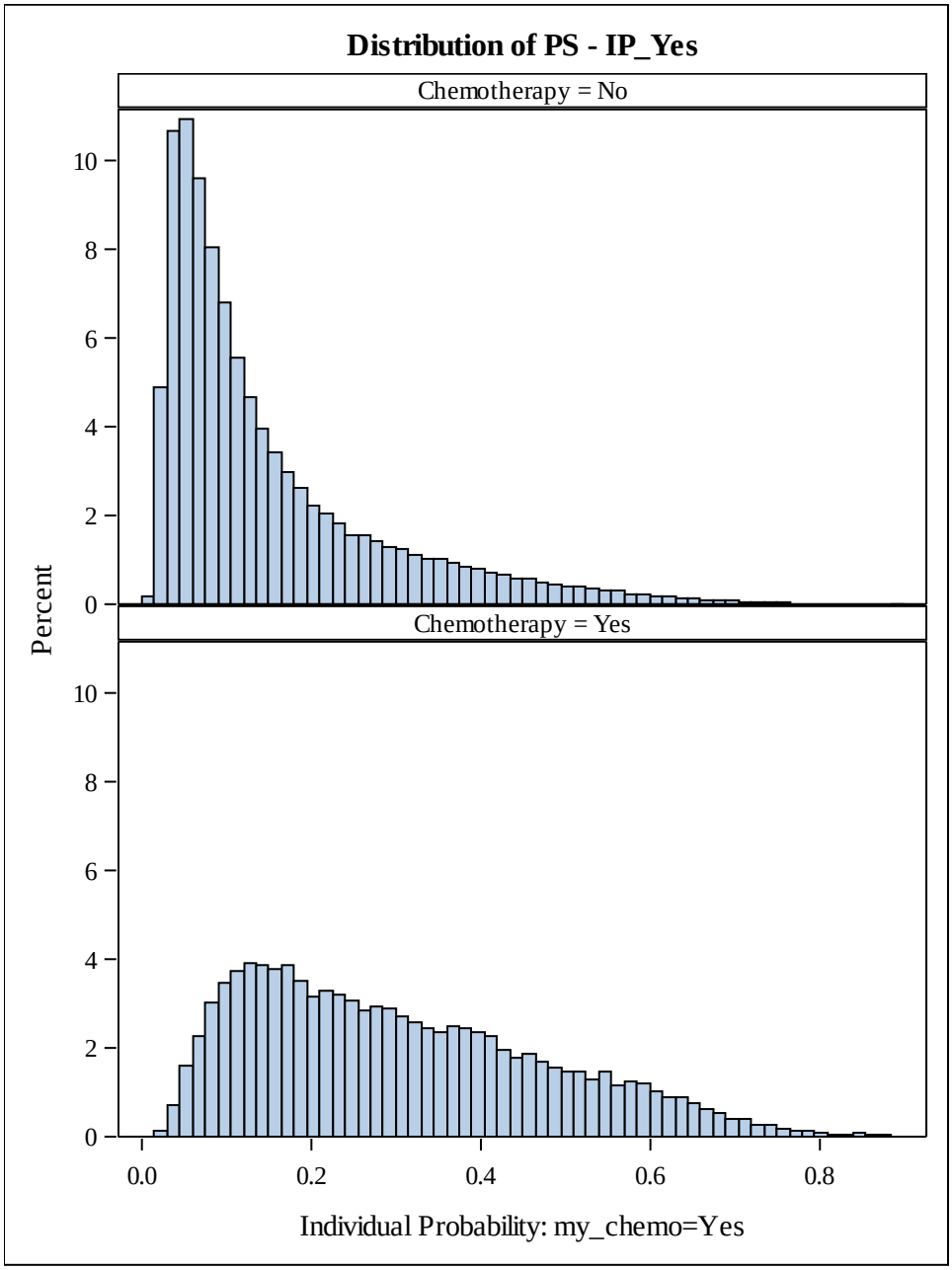


Figure 2. Distribution of propensity score matching

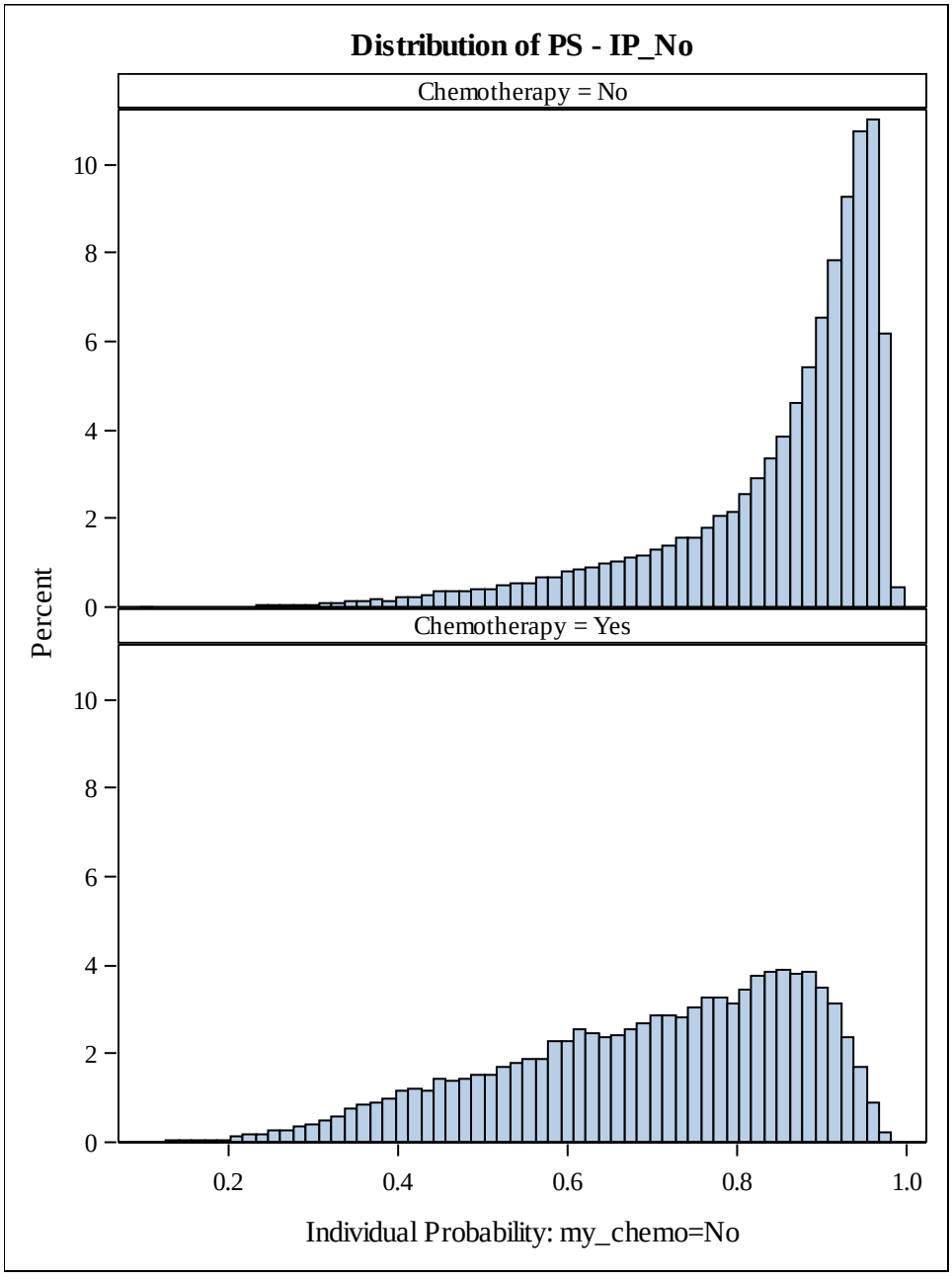


Figure3. Distribution of propensity score matching